

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

01-OCT-2003 (TEMBLrel. 25, Last annotation update)
ATP synthase F0 subunit 8.
Name=ATP8;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=22406325; PubMed=12509511;
RA Mishmar D., Ruiz-Pesini E., Golik P., Macaulay V., Clark A.G.,
RS Hosseini S., Brandon M., Easley K., Chen B., Brown M.D.,
RA Sukernik R.I., Olckers A., Wallace D.C.;
RT "Natural selection shaped regional mtDNA variation in humans.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:171-176(2003).
CC -!- FUNCTION: This is one of the chains of the nonenzymatic component
of the mitochondrial ATPase complex (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
H(+)(Out)
CC -!- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -!- SIMILARITY: Belongs to the ATPase protein 8 family.
DR EMBL; AV195782; AAC8765.1; -.
DR GO; GO:0005739; C:mitochochdron; IEA.
DR GO; GO:0016469; C:proton-transporing two-sector ATPase complex; IEA.
DR GO; GO:0046933; F:hydrogen-transporing ATP synthase activity. . ; IEA.
DR GO; GO:0046961; F:hydrogen-transporing ATPase activity, rota. . ; IEA.
DR GO; GO:0015986; P:ATP synthesis coupled proton transpor; IEA.
DR GO; GO:0015992; P:proton transport; IEA.
DR InterPro; IPR01421; ATPases mit.
DR Pfam; PF00895; ATP-synt 8; 1.
KW CF(O); Hydrogen ion transport; Ion transport; Mitochondrion;
Transmembrane; Transport
SQ SEQUENCE 68 AA; 8012 MW; COD7B513AD7692B CRC64;

Query Match 68.3%; Score 102.5; DB 2; Length 68;
Best Local Similarity 69.0%; Pred. No. 2.e-06;
Matches 20; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

QY 2 RLNTTWPITIPILLTLFLTN-RLITT 29 :|||||:|||||::|||:::
DB 3 QLNTTWPTIPEWLTLLFIOLKMENT 31

RESULT 10
Q85KZ3 PRELIMINARY; PRT; 68 AA.
AC Q85KZ3
DT 01-JUN-2003 (TEMBLrel. 24, Created)
DI 01-JUN-2003 (TEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE ATP synthase F0 subunit 8.
GN Name=ATP8;
GS OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=22406325; PubMed=12509511;
RA Mishmar D., Ruiz-Pesini E., Golik P., Macaulay V., Clark A.G.,
RS Hosseini S., Brandon M., Easley K., Chen B., Brown M.D.,
RA Sukernik R.I., Olckers A., Wallace D.C.;
RT "Natural selection shaped regional mtDNA variation in humans.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:171-176(2003).
CC -!- FUNCTION: This is one of the chains of the nonenzymatic component
of the mitochondrial ATPase complex (By similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+)(In) = ADP + phosphate +
H(+)(Out).
CC -!- SUBCELLULAR LOCATION: Membrane-bound (By similarity).

```

CC -!- SIMILARITY: Belongs to the ATPase protein 8 family.
DR EMBL; AY195756; AAC88427.1; .
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.
DR GO; GO:0046933; F:hydrogen-transporting ATP synthase activity, rota. . ; IEA.
DR GO; GO:0046961; F:hydrogen-transporting ATPase activity, rota. . ; IEA.
DR GO; GO:0046961; F:hydrogen-transporting ATPase activity, rota. . ; IEA.
DR GO; GO:0015986; P:ATP synthesis coupled proton transport; IEA.
DR GO; GO:0015992; P:proton transport; IEA.
DR InterPro; IPR001421; ATPase8_mit.
DR Pfam; PF00895; ATP-synt_8; 1.
DR CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
KW Transmembrane; Transport.
SQ SEQUENCE 68 AA; 8008 MW; 6C756A113AD76936 CRC64;

Query Match 68.3%; Score 102.5; DB 2; Length 68;
Best Local Similarity 69.0%; Pred. No. 2.2e-06;
Matches 20; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

QY 2 RLNTTWPMTIIPILLTLFLITN-RLITT 29
Db 3 QLNTTWPMTIIPMTLLTLFLITQLKMLNT 31

RESULT 11
ID Q8HNQ6 PRELIMINARY; PRT; 68 AA.
AC Q8HNQ6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE ATP synthase 8.
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22062553; PubMed=12022039;
RA Silva W.A. Jr., Bonatto S.L., Holanda A.J., Ribeiro-Dos-Santos A.K.,
RA Paixao B.M., Goldman G.H., Abe-Sandes K., Rodriguez-Delfin L.,
RA Batdosa M., Peco-Larson M.L., Petzl-Erler M.L., Valente V.,
RA Santos S.E., Zago M.A.;
RT "Mitochondrial genome diversity of Native Americans supports a single
RT early entry of founder populations into America.";
RL Am. J. Hum. Genet. 71:187-192(2002).
CC -!- FUNCTION: This is one of the chains of the nonenzymatic component
CC (CF(0) subunit) of the mitochondrial ATPase complex (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -!- SIMILARITY: Belongs to the ATPase protein 8 family.
DR EMBL; AF346999; AAK17679.1; .
DR EMBL; AF346998; AAK17666.1; .
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.
DR GO; GO:0046933; F:hydrogen-transporting ATP synthase activity, rota. . ; IEA.
DR GO; GO:0046961; F:hydrogen-transporting ATPase activity, rota. . ; IEA.
DR GO; GO:0015986; P:ATP synthesis coupled proton transport; IEA.
DR GO; GO:0015992; P:proton transport; IEA.
DR InterPro; IPR001421; ATPase8_mit.
DR Pfam; PF00895; ATP-synt_8; 1.
DR CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
KW Transmembrane; Transport.
SQ SEQUENCE 68 AA; 8002 MW; 7D307E513AD5B936 CRC64;

Query Match 68.3%; Score 102.5; DB 2; Length 68;
Best Local Similarity 69.0%; Pred. No. 2.2e-06;
Matches 20; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

QY 2 RLNTTWPMTIIPILLTLFLITN-RLITT 29
Db 3 QLNTTWPMTIIPMTLLTLFLITQLKMLNT 31

RESULT 12
ID Q94P44 PRELIMINARY; PRT; 68 AA.
AC Q94P44;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE ATP synthase 8.
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21012010; PubMed=11130070;
RA Ingman M., Kaessmann H., Paabo S., Gyllenstein U.;
RT "Mitochondrial genome variation and the origin of modern humans.";
RL Nature 408:708-713(2000).
CC -!- FUNCTION: This is one of the chains of the nonenzymatic component
CC (CF(0) subunit) of the mitochondrial ATPase complex (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
CC H(+) (Out).
CC -!- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
CC -!- SIMILARITY: Belongs to the ATPase protein 8 family.
DR EMBL; AF346999; AAK17679.1; .
DR EMBL; AF346998; AAK17666.1; .
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0016469; C:proton-transporting two-sector ATPase complex; IEA.
DR GO; GO:0046933; F:hydrogen-transporting ATP synthase activity, rota. . ; IEA.
DR GO; GO:0046961; F:hydrogen-transporting ATPase activity, rota. . ; IEA.
DR GO; GO:0015986; P:ATP synthesis coupled proton transport; IEA.
DR GO; GO:0015992; P:proton transport; IEA.
DR InterPro; IPR001421; ATPase8_mit.
DR Pfam; PF00895; ATP-synt_8; 1.
DR CF(0); Hydrogen ion transport; Ion transport; Mitochondrion;
KW Transmembrane; Transport.
SQ SEQUENCE 68 AA; 7965 MW; 7D307E512A3967C6 CRC64;

Query Match 68.3%; Score 102.5; DB 2; Length 68;
Best Local Similarity 69.0%; Pred. No. 2.2e-06;
Matches 20; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

QY 2 RLNTTWPMTIIPILLTLFLITN-RLITT 29
Db 3 QLNTTWPMTIIPMTLLTLFLITQLKMLNT 31

RESULT 13
ID AAL54388 PRELIMINARY; PRT; 68 AA.
AC AAL54388;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE ATP synthase 8.
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11553319;
Maca-Meyer N., Gonzalez A.M., Larruga J.M., Flores C., Cabrera V.M.;

```



```

RT "Major genomic mitochondrial lineages delineate early human
RL expansions.";
RL BMC Genet. 2:13-13(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Cabrera V.M., Maca-Meyer N., Gonzalez A.M., Larruga J.M., Flores C.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF381981; AAL54388.1; -.
KW Mitochondrion.
SQ SEQUENCE 68 AA; 7992 MW; 7D307E513AD76936 CRC64;

Query Match 68.3%; Score 102.5; DB 2; Length 68;
Best Local Similarity 69.0%; Pred. No. 2.2e-06;
Matches 20; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

Qy 2 RLNTVWPTIITPILLTFLITN-RLITT 29
:|||||:|||||:|||||:|||||:|||||:
Db 3 QLNTVWPTMITPMLLTFLITQLKMLNT 31
:|||||:|||||:|||||:|||||:|||||:

RESULT 14
AAL54414 PRELIMINARY; PRT; 68 AA.
ID AAL54414
AC AAL54414;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE ATP synthase 8.
GN ATP8.
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Maca-Meyer N., Gonzalez A.M., Larruga J.M., Flores C., Cabrera V.M.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF381983; AAL54414.1; -.
KW Mitochondrion.
SQ SEQUENCE 68 AA; 7992 MW; 7D307E513AD76936 CRC64;

Query Match 68.3%; Score 102.5; DB 2; Length 68;
Best Local Similarity 69.0%; Pred. No. 2.2e-06;
Matches 20; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

Qy 2 RLNTVWPTIITPILLTFLITN-RLITT 29
:|||||:|||||:|||||:|||||:|||||:
Db 3 QLNTVWPTMITPMLLTFLITQLKMLNT 31
:|||||:|||||:|||||:|||||:|||||:

RESULT 15
AAL54427 PRELIMINARY; PRT; 68 AA.
ID AAL54427
AC AAL54427;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE ATP synthase 8.
GN ATP8.
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX PubMed=11553319;
RA Maca-Meyer N., Gonzalez A.M., Larruga J.M., Flores C., Cabrera V.M.;
RT "Major genomic mitochondrial lineages delineate early human
RT expansions.";
RL BMC Genet. 2:13-13(2001).
RN [2]
RP SEQUENCE FROM N.A.
RA Cabrera V.M., Maca-Meyer N., Gonzalez A.M., Larruga J.M., Flores C.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF381984; AAL54427.1; -.
KW Mitochondrion.
SQ SEQUENCE 68 AA; 7992 MW; 7D307E513AD76936 CRC64;

Query Match 68.3%; Score 102.5; DB 2; Length 68;
Best Local Similarity 69.0%; Pred. No. 2.2e-06;
Matches 20; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

Qy 2 RLNTVWPTIITPILLTFLITN-RLITT 29
:|||||:|||||:|||||:|||||:|||||:
Db 3 QLNTVWPTMITPMLLTFLITQLKMLNT 31
:|||||:|||||:|||||:|||||:|||||:

Search completed: November 10, 2004, 13:38:23
Job time : 36.6321 secs

```

This page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:32:37 ; Search time 8.981 Seconds
(without alignments)
191.991 Million cell updates/sec

Title: US-10-092-750-28
Perfect score: 127
Sequence: 1 TLVLTALAVTFGLTALDLYPT 26

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PTCUS.COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	88.2	603	3	US-09-097-889-23
2	112	88.2	603	4	US-09-098-079-23
3	48.5	38.2	57	1	US-08-137-614A-16
4	48.5	38.2	686	2	US-08-768-301-4
5	48	37.8	340	4	US-09-328-352-4501
6	48	37.8	525	4	US-09-543-681A-5886
7	48	37.8	625	4	US-09-583-545-2
8	48	37.8	625	4	US-09-617-594A-6
9	47	37.0	238	4	US-09-452-239-28
10	47	37.0	480	2	US-08-468-812-5
11	47	37.0	480	3	US-08-590-563-5
12	47	37.0	480	4	US-09-770-621-5
13	47	37.0	480	4	US-09-235-832-5
14	47	37.0	492	2	US-08-468-812-4
15	47	37.0	492	2	US-08-468-812-7
16	47	37.0	492	3	US-08-590-563-4
17	47	37.0	492	3	US-08-590-563-7
18	47	37.0	492	4	US-09-770-621-4
19	47	37.0	492	4	US-09-770-621-7
20	47	37.0	492	4	US-09-235-832-4
21	47	37.0	492	4	US-09-235-832-7
22	45	35.4	67	4	US-09-270-767-35719
23	45	35.4	67	4	US-09-270-767-50936
24	45	35.4	137	4	US-09-452-239-34
25	45	35.4	171	4	US-09-248-796A-20109
26	45	35.4	425	4	US-09-489-039A-8439
27	45	35.4	426	4	US-09-583-110-4783

28 45 35.4 895 3 US-08-827-962-19 Sequence 19, Appl
29 45 35.4 895 3 US-08-827-962-21 Sequence 21, Appl
30 45 35.4 898 2 US-08-808-982-5 Sequence 5, Appl
31 45 35.4 898 3 US-09-306-902A-5 Sequence 5, Appl
32 45 35.4 1162 3 US-08-827-962-15 Sequence 15, Appl
33 45 35.4 1162 3 US-08-827-962-20 Sequence 20, Appl
34 45 35.4 1162 3 US-08-803-345-1 Sequence 1, Appl
35 44 34.6 446 4 US-09-489-039A-14283 Sequence 14283, A
36 44 34.6 840 4 US-09-252-991A-19561 Sequence 19561, A
37 43.5 34.3 391 3 US-07-741-453A-2 Sequence 2, Appl
38 43.5 34.3 764 3 US-07-757-342D-5 Sequence 5, Appl
39 43.5 34.3 764 3 US-07-741-453A-29 Sequence 29, Appl
40 43.5 34.3 764 3 US-07-741-453A-54 Sequence 54, Appl
41 43.5 34.3 764 3 US-07-741-453A-59 Sequence 59, Appl
42 43.5 34.3 764 3 US-07-741-453A-60 Sequence 60, Appl
43 43.5 34.3 764 3 US-07-741-453A-61 Sequence 61, Appl
44 43.5 34.3 764 4 US-09-461-657B-5 Sequence 5, Appl
45 43.5 34.3 764 4 US-09-186-350A-53 Sequence 53, Appl

ALIGNMENTS

RESULT 1
US-09-097-889-23 Application US/09097889
; Sequence 23, Patent No. 6218117
; GENERAL INFORMATION:
; APPLICANT: Hevinstadt, Corrina
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Davis, Robert E.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING
; TITLE OF INVENTION: AGENTS THAT QUANTITATIVELY ALTER DETECTABLE
; TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA: MITOCHONDRIAL DNA RATIOS
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/097,889
; FILING DATE: 15-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Roseman Ph.D., Stephen J.
; REGISTRATION NUMBER: 43,058
; REFERENCE/DOCKET NUMBER: 660088.417
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 603 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-097-889-23

Query Match 88.2%; Score 112; DB 3; Length 603;
Best Local Similarity 96.0%; Pred. No. 4.3e-09;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 LVLTALAVTFGLTALDLYPT 26
DB 484 LVLTALAVTFGLTALDLYPT 508


```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; US-08-768-301-4

Query Match      38.2%; Score 48.5; DB 2; Length 686;
Best Local Similarity 52.2%; Pred. No. 33;
Matches 12; Conservative 2; Mismatches 6; Indels 3; Gaps 1;

QY 7 TALAVTFGLTALDNL---NYPT 26
DB 430 TVLTMTFLGLEARTDLPKVSYP 452

RESULT 5
US-09-328-352-4501
; Sequence 4501, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 4501
; LENGTH: 340
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
; US-09-328-352-4501

Query Match      37.8%; Score 48; DB 4; Length 340;
Best Local Similarity 47.8%; Pred. No. 17;
Matches 11; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDNLNY 24
DB 34 IYLVICLAVCFGLNLSGLDFDY 56

RESULT 6
US-09-543-681A-5886
; Sequence 5886, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5886
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; US-09-543-681A-5886

Query Match      37.8%; Score 48; DB 4; Length 525;
Best Local Similarity 45.5%; Pred. No. 29;
Matches 10; Conservative 5; Mismatches 5; Indels 2; Gaps 1;

QY 5 KLTALAVTFGLLT---TALDNLNY 24
DB 129 KMTTLAIVFAGIMSNTAAELGY 150

```

```

RESULT 7
US-09-583-545-2
; Sequence 2, Application US/09583545
; Patent No. 6437883
; GENERAL INFORMATION:
; APPLICANT: BUELOT et al.
; TITLE OF INVENTION: PORCINE CIRCOVIRUS RECOMBINANT POXVIRUS VACCINE
; FILE REFERENCE: 454313-2511.1
; CURRENT APPLICATION NUMBER: US/09/583,545
; CURRENT FILING DATE: 2000-06-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Canarypox virus
; US-09-583-545-2

Query Match      37.8%; Score 48; DB 4; Length 625;
Best Local Similarity 45.5%; Pred. No. 35;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDNL 23
DB 496 LYSKDISAKRFGLFTSVGIN 517

```

```

RESULT 8
US-09-617-594A-6
; Sequence 6, Application US/09617594A
; Patent No. 6541458
; GENERAL INFORMATION:
; APPLICANT: Audonnet, et al.
; TITLE OF INVENTION: FELINE CALICIVIRUS GENES AND VACCINES IN PARTICULAR RECOMBINANT V
; FILE REFERENCE: 454313-3151.1
; CURRENT APPLICATION NUMBER: US/09/617,594A
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/193,332
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: France 00 01761
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: France 99 09421
; PRIOR FILING DATE: 1999-07-16
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 625
; TYPE: PRT
; ORGANISM: canarypox virus
; US-09-617-594A-6

Query Match      37.8%; Score 48; DB 4; Length 625;
Best Local Similarity 45.5%; Pred. No. 35;
Matches 10; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDNL 23
DB 496 LYSKDISAKRFGLFTSVGIN 517

```

```

RESULT 9
US-09-452-239-28
; Sequence 28, Application US/09452239
; Patent No. 6465229
; GENERAL INFORMATION:
; APPLICANT: Rafalski, Antoni J.
; APPLICANT: Fader, Gary M.
; APPLICANT: Caloon, Rebecca E.
; TITLE OF INVENTION: Plant Caffeoyl-CoA O-Methyltransferase
; FILE REFERENCE: BB1284 US NA
; CURRENT APPLICATION NUMBER: US/09/452,239

```

; CURRENT FILING DATE: 1999-12-01
; EARLIER APPLICATION NUMBER: 60/110,594
; EARLIER FILING DATE: 1998-December-02
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 28
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Glycine max
US-09-452-239-28

Query Match 37.0%; Score 47; DB 4; Length 238;
Best Local Similarity 47.8%; Pred. No. 16;
Matches 11; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 4 LKLTALAVTFLGLLTALDNLNPT 26
||| : : : : :
Db 86 LLLTALNIPHDGKITAIDNRKT 108

RESULT 10
US-08-468-812-5
; Sequence 5, Application US/08468812
; Patent No. 5935836
; GENERAL INFORMATION:
; APPLICANT: Vehmaanper, Jari
; APPLICANT: M ntyl Arja
; APPLICANT: Fagerstr m, Richard
; APPLICANT: Lantto, Raija
; APPLICANT: Paloheimo, Marja
; APPLICANT: Suominen, Pirkko
; APPLICANT: Lahtinen, Tarja
; APPLICANT: Kristo, Paula
; TITLE OF INVENTION: Actinomadura Xylanase Sequences and Methods
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,812
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/332,412
FILING DATE: 31-OCT-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/282,001
FILING DATE: 29-JUL-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Bugaisky, Larry B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 1050.0340002
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear

; MOLECULE TYPE: peptide
; POSITION IN GENOME:
; CHROMOSOME/SEGMENT: AM50
US-08-468-812-5

Query Match 37.0%; Score 47; DB 2; Length 480;
Best Local Similarity 41.7%; Pred. No. 37;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFLGLLTALDNLNYP 25
||| : : : : :
Db 18 LYRALAAATVSVGVVVTALTVTOP 41

RESULT 11
US-08-590-563-5
; Sequence 5, Application US/08590563
; Patent No. 6300114
; GENERAL INFORMATION:
; APPLICANT: M ntyl, Arja
; APPLICANT: Vehmaanper, Jari
; APPLICANT: Fagerstr m, Richard
; APPLICANT: Lantto, Raija
; APPLICANT: Paloheimo, Marja
; APPLICANT: Suominen, Pirkko
; APPLICANT: Lahtinen, Tarja
; TITLE OF INVENTION: Production and Secretion of Proteins of
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
; STREET: 1100 New York Ave., N.W. Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/590,563
FILING DATE: 26-JAN-1996
CLASSIFICATION: 536

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/468,812
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/332,412
FILING DATE: 31-OCT-1994
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/282,001
FILING DATE: 29-JUL-1994
CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:
NAME: Bugaisky, Lawrence B.
REGISTRATION NUMBER: 35,086
REFERENCE/DOCKET NUMBER: 1050.0340003
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
POSITION IN GENOME: AM50
CHROMOSOME/SEGMENT: AM50
US-08-590-563-5

Query Match 37.0%; Score 47; DB 4; Length 480;
Best Local Similarity 41.7%; Pred. No. 37;
Matches 10; Conservative 5; Mismatches 9; Indels

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 16:36:12 ; Search time 27.9309 Seconds
(without alignments)
328.807 Million cell updates/sec

Title: US-10-092-750-28

Perfect score: 127

Sequence: 1 TLVLTALAVTFGLTALDNLNPT 26

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	127	100.0	26	14 US-10-092-750-28	Sequence 28, Appl
2	112	88.2	603	9 US-09-098-079-23	Sequence 23, Appl
3	112	88.2	603	16 US-10-408-765A-1487	Sequence 1487, Ap
4	112	88.2	603	16 US-10-408-765A-1489	Sequence 1489, Ap
5	112	88.2	603	16 US-10-408-765A-2681	Sequence 2681, Ap
6	61	48.0	610	14 US-10-205-219-71	Sequence 71, Appl
7	53	41.7	299	9 US-09-825-882-16	Sequence 16, Appl
8	53	41.7	299	14 US-10-017-161-1832	Sequence 1832, Ap
9	53	41.7	299	14 US-10-292-798-1488	Sequence 1488, Ap
10	53	41.7	299	15 US-10-343-650A-694	Sequence 694, Ap
11	53	41.7	299	17 US-10-724-208-16	Sequence 16, Appl
12	53	41.7	299	17 US-10-770-127-187	Sequence 187, Appl
13	52	40.9	881	16 US-10-437-963-198410	Sequence 198410,

14	50	39.4	95	17	US-10-425-115-323215	Sequence 323215,
15	50	39.4	112	15	US-10-424-599-241860	Sequence 241860,
16	50	39.4	214	15	US-10-335-977-4897	Sequence 4897, Ap
17	50	39.4	230	15	US-10-335-977-4898	Sequence 4898, Ap
18	50	39.4	340	15	US-10-425-114-52996	Sequence 52996, A
19	50	39.4	423	17	US-10-425-115-187565	Sequence 187565,
20	49	38.6	157	14	US-10-243-552-635	Sequence 635, App
21	49	38.6	219	15	US-10-424-599-146138	Sequence 146138,
22	49	38.6	275	16	US-10-474-776-330	Sequence 330, Appl
23	49	38.6	309	9	US-09-825-882-20	Sequence 20, Appl
24	49	38.6	309	14	US-10-017-161-1932	Sequence 1932, Ap
25	49	38.6	309	14	US-10-292-798-1584	Sequence 1584, Ap
26	49	38.6	309	17	US-10-724-208-20	Sequence 20, Appl
27	49	38.6	309	17	US-10-770-127-191	Sequence 191, Appl
28	49	38.6	855	15	US-10-282-122A-64478	Sequence 64478, A
29	48	37.8	218	13	US-10-062-254-56	Sequence 56, Appl
30	48	37.8	220	16	US-10-437-963-134834	Sequence 134834,
31	48	37.8	299	10	US-09-510-332-53	Sequence 53, Appl
32	48	37.8	299	17	US-10-770-127-53	Sequence 53, Appl
33	48	37.8	309	9	US-09-825-882-8	Sequence 8, Appl
34	48	37.8	309	9	US-09-825-882-12	Sequence 12, Appl
35	48	37.8	309	14	US-10-219-834-18	Sequence 18, Appl
36	48	37.8	309	14	US-10-017-161-1836	Sequence 1836, Ap
37	48	37.8	309	14	US-10-017-161-1838	Sequence 1838, Ap
38	48	37.8	309	14	US-10-191-088-5	Sequence 5, Appl
39	48	37.8	309	14	US-10-292-798-1492	Sequence 1492, Ap
40	48	37.8	309	14	US-10-292-798-1494	Sequence 1494, Ap
41	48	37.8	309	15	US-10-343-650A-682	Sequence 682, App
42	48	37.8	309	15	US-10-343-650A-692	Sequence 692, App
43	48	37.8	309	17	US-10-724-208-8	Sequence 8, Appl
44	48	37.8	309	17	US-10-724-208-12	Sequence 12, Appl
45	48	37.8	309	17	US-10-770-127-179	Sequence 179, App

ALIGNMENTS

RESULT 1
US-10-092-750-28
; Sequence 28, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-28

Query Match 100.0%; Score 127; DB 14; Length 26;
Best Local Similarity 100.0%; Pred. No. 5.3e-12;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TLVLTALAVTFGLTALDNLNPT 26

Db 1 TLVLTALAVTFGLTALDNLNPT 26

RESULT 2

US-09-098-079-23
; Sequence 23, Application US/09098079
; Patent No. US20020064773A1
; GENERAL INFORMATION:

/ APPLICANT: Herrnstadt, Corrina
 / APPLICANT: Ghosh, Soumitra S.
 / APPLICANT: Clevenger, William
 / APPLICANT: Fahy, Eoin F.
 / APPLICANT: Davis, Robert E.
 / TITLE OF INVENTION: DIAGNOSTIC METHOD BASED ON QUANTIFICATION OF
 / TITLE OF INVENTION: EXTRAMITOCHONDRIAL DNA
 / NUMBER OF SEQUENCES: 26
 / CORRESPONDENCE ADDRESS:
 / ADDRESS: SEED and BERRY LLP
 / CITY: Seattle
 / STATE: Washington
 / COUNTRY: USA
 / ZIP: 98104
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: PatentIn Release #1.0, Version #1.30
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/09/098,079
 / FILING DATE: 15-JUN-1998
 / CLASSIFICATION: 435
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Roserman Ph.D., Stephen J.
 / REGISTRATION NUMBER: 43,058
 / REFERENCE/DOCKET NUMBER: 860088.416
 / TELEPHONE: (206) 622-4900
 / TELEFAX: (206) 682-6031
 / INFORMATION FOR SEQ ID NO: 23:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 603 amino acids
 / TYPE: amino acid
 / STRANDEDNESS:
 / TOPOLOGY: linear
 /

```
Query Match      88.2%; Score 112; DB 9; Length 603;
Best Local Similarity 96.0%; Pred. No. 3.4e-08;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

Qy 2 LYLKLTALAVTFELGLTALDLYPT 26
| | | | | | | | | | | | | |
Db 484 LYLKLTALAVTFELGLTALDLYLT 508

```

RESULT 3
US-10-408-765A-1487
; Sequence 1487, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fany, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1487
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1487

```

```

Query Match      88.2%; Score 112; DB 16; Length 603;
Best Local Similarity 96.0%; Pred. No. 3.4e-08;
Matches 24; Conservative 0; Mismatches 1; Indels

QY    2 LYLKLTALAVTFLGLTALDLYNPT 26
      |||||
DB     484 LYLKLTALAVTFLGLTALDLYNLT 508

```

```

RESULT 4
US-10-408-765A-1489
; Sequence 1489, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1489
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-1489

```

Query Match 88.2%; Score 112; DB 16; Length 603;
Best Local Similarity 96.0%; Pred. No. 3.4e-08;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 2 LYIKLTALAVTFGLLTALDLYPT 26
484 LYIKLTALAVTFGLLTALDLYLT 508

```

RESULT 5
US-10-408-765A-3681
; Sequence 2681, Application US/10408765A
; Publication No. US2004010187A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale B.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2681
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2681

```

Query Match 88.2%; Score 112; DB 16; Length 603;
Best Local Similarity 96.0%; Pred. No. 3.4e-08;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFLGLLTALDLNYPT 26

```
Db 484 LYLKLTALAVTFLGLLTALDNLNLT 508
|||||
RESULT 6
; Sequence 71, Application US/10205219
; Publication No. US20030134803A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brooksbank, Robert
; APPLICANT: Pincock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018200
; CURRENT APPLICATION NUMBER: US/10/205,219
; CURRENT FILING DATE: 2002-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 71
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: NADH subunit 5
US-10-205-219-71

Query Match 48.0%; Score 61; DB 14; Length 610;
Best Local Similarity 57.1%; Pred. No. 1.6;
Matches 12; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 3 YLKLTALAVTFLGLLTALDNLN 23
:|||||:
Db 485 HLKMTALLITLFGAIALELN 505
|||||

RESULT 7
US-09-825-982-16
; Sequence 16, Application US/09825882
; Patent No. US20020094551A1
; GENERAL INFORMATION:
; APPLICANT: ADLER, JON ELLIOT
; TITLE OF INVENTION: T2R TASTE RECEPTORS AND GENES ENCODING SAME
; FILE REFERENCE: 078003/0279152/EXT
; CURRENT APPLICATION NUMBER: US/09/825,882
; CURRENT FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/195,532
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/247,014
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 16
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-825-982-16

Query Match 41.7%; Score 53; DB 9; Length 299;
Best Local Similarity 68.4%; Pred. No. 12;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 6 LTALAVTFLGLLTALDNLN 24
|||||
Db 48 LTALAVSRIGLLWALLLNW 66
|||||

RESULT 8
US-10-017-161-1832
; Sequence 1832, Application US/10017161
```

```
; PRIOR APPLICATION NUMBER: JP 2000/237818
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: JP 2001/34434
; PRIOR FILING DATE: 2001-02-13
; NUMBER OF SEQ ID NOS: 694
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 694
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-343-650A-694

Query Match          41.7%; Score 53; DB 15; Length 299;
Best Local Similarity 68.4%; Pred. No. 12;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 6 LTALAVTFLGLTALDNLN 24
DB 48 LTALAVSRIGLLWALLLNW 66

RESULT 11
US-10-724-208-16
; Sequence 16, Application US/10724208
; Publication No. US20040209313A1
; GENERAL INFORMATION:
; APPLICANT: ADLER, JON ELLIOT
; TITLE OF INVENTION: T2R TASTE RECEPTORS AND GENES ENCODING SAME
; FILE REFERENCE: 078003/0279152/RXT
; CURRENT APPLICATION NUMBER: US/10/724,208
; CURRENT FILING DATE: 2003-12-01
; PRIOR APPLICATION NUMBER: 60/195,532
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/247,014
; PRIOR FILING DATE: 2000-11-13
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-724-208-16

Query Match          41.7%; Score 53; DB 17; Length 299;
Best Local Similarity 68.4%; Pred. No. 12;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 6 LTALAVTFLGLTALDNLN 24
DB 48 LTALAVSRIGLLWALLLNW 66

RESULT 12
US-10-770-127-187
; Sequence 187, Application US/10770127
; Publication No. US20040214239A1
; GENERAL INFORMATION:
; APPLICANT: SERVANT, GUY
; APPLICANT: OZECK, MARK
; APPLICANT: BRUST, PAUL
; APPLICANT: XU, HONG
; TITLE OF INVENTION: FUNCTIONAL COUPLING OF T1RS AND T2RS BY GI PROTEINS
; TITLE OF INVENTION: AND CELL-BASED ASSAYS FOR THE IDENTIFICATION OF T1R
; TITLE OF INVENTION: AND T2R MODULATORS
; FILE REFERENCE: 100337.54281US
; CURRENT APPLICATION NUMBER: US/10/770,127
; CURRENT FILING DATE: 2004-02-03
; PRIOR APPLICATION NUMBER: 60/444,172
; PRIOR FILING DATE: 2003-02-03
; PRIOR APPLICATION NUMBER: 60/457,318
; PRIOR FILING DATE: 2003-03-26
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 187
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-770-127-187

Query Match          41.7%; Score 53; DB 17; Length 299;
Best Local Similarity 68.4%; Pred. No. 12;
Matches 13; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 6 LTALAVTFLGLTALDNLN 24
DB 48 LTALAVSRIGLLWALLLNW 66

RESULT 13
US-10-437-963-198410
; Sequence 198410, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 198410
; LENGTH: 881
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_94072C.1.pep
US-10-437-963-198410

Query Match          40.9%; Score 52; DB 16; Length 881;
Best Local Similarity 61.1%; Pred. No. 56;
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 YLKLTAATVFLGLTAL 20
DB 700 YLQSSAMVFLGLLFAI 717

RESULT 14
US-10-425-115-323215
; Sequence 323215, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 323215
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(95)
```

```
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_57841C.1.pep
US-10-425-115-323215

Query Match      39.4%; Score 50; DB 17; Length 95;
Best Local Similarity 52.4%; Pred. No. 9;
Matches 11; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY 4 LKLTALAVTFGLLTALDINY 24
   |||:|:|:|:|:|:|:|:|:|
Db 18 LLLSTLTSGGLLSALSINY 38

RESULT 15
US-10-424-599-241860
; Sequence 241860, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 241860
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_60427C.1.pep
US-10-424-599-241860

Query Match      39.4%; Score 50; DB 15; Length 112;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 9; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 8 ALAVTFGLLTALDINYP 25
   |||:|:|:|:|:|:|:|
Db 91 AISSTFLQVLTSLISYP 108

Search completed: November 11, 2004, 01:28:10
Job time : 28.9809 secs
```

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:29:32 ; Search time 5.79275 Seconds
(without alignments)
431.857 Million cell updates/sec

Title: US-10-092-750-28
Perfect score: 127
Sequence: 1 TLYIKLTALAVTFGLTALDLYPT 26

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	112	89.2	603	1 DNRHNS	NADH2 dehydrogenas
2	109	85.8	603	2 T14149	NADH2 dehydrogenas
3	102	80.3	603	2 B59154	NADH2 dehydrogenas
4	102	80.3	603	2 T14036	NADH2 dehydrogenas
5	100	78.7	603	2 T14199	NADH2 dehydrogenas
6	86	67.7	603	2 T14145	NADH2 dehydrogenas
7	81	63.8	603	2 T11843	NADH2 dehydrogenas
8	73.5	57.9	606	2 T11503	NADH2 dehydrogenas
9	73.5	57.9	609	2 S26161	NADH2 dehydrogenas
10	72.5	57.1	606	2 T11412	NADH2 dehydrogenas
11	72	56.7	606	2 T10982	NADH2 dehydrogenas
12	70	55.1	603	2 T11399	NADH2 dehydrogenas
13	69.5	54.7	606	2 T11060	NADH2 dehydrogenas
14	69	54.3	601	2 T11451	NADH2 dehydrogenas
15	69	54.3	604	2 T11867	NADH2 dehydrogenas
16	68	53.5	606	2 CS8851	NADH2 dehydrogenas
17	68	53.5	613	2 T11464	NADH2 dehydrogenas
18	67	52.8	609	2 S41845	NADH2 dehydrogenas
19	67	52.8	612	2 T11830	NADH2 dehydrogenas
20	66	52.0	606	2 T11150	NADH2 dehydrogenas
21	66	52.0	606	2 T11373	NADH2 dehydrogenas
22	65.5	51.6	609	2 T11310	NADH2 dehydrogenas
23	65.5	51.6	611	2 T11297	NADH2 dehydrogenas
24	64	50.4	603	2 T11490	NADH2 dehydrogenas
25	64	50.4	606	2 S41830	NADH2 dehydrogenas
26	62	48.8	604	2 T11347	NADH2 dehydrogenas
27	62	48.8	606	2 T11334	NADH2 dehydrogenas
28	62	48.8	645	2 T11137	NADH2 dehydrogenas
29	61	48.0	606	1 OXBOSM	NADH2 dehydrogenas

RESULT 1

DNRHNS

NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - human mitochondrion
N;Alternate names: NADH-ubiquinone oxidoreductase chain 5
C;Species: mitochondrion Homo sapiens (man)
C;Date: 22-May-1981 #sequence revision 23-Oct-1981 #text_change 09-Jul-2004
C;Accession: A00446; C00435; I80239

R;Anderson, S.; Bankier, A.T.; Barrell, B.G.; de Bruijn, M.H.L.; Coulson, A.R.; Drouin, J.

Nature 290, 457-465, 1981

A;Title: Sequence and organization of the human mitochondrial genome.

A;Reference number: A00151; MUID:81173052; PMID:7219534

A;Accession: A00446

A;Molecule type: DNA

A;Residues: 1-603 <AND>

A;Cross-references: UNIPROT:P03915; GB:J01415; NID:g1944628; PIDN:AAB58953.1; PID:320523

R;Brown, W.M.; Prager, E.M.; Wang, A.; Wilson, A.C.

J. Mol. Evol. 18, 225-239, 1982

A;Title: Mitochondrial DNA sequences of primates: tempo and mode of evolution.

A;Reference number: A00435; MUID:82242101; PMID:6284948

A;Accession: C00435

A;Molecule type: DNA

A;Residues: 1-79 <BRO>

A;Cross-references: GB:L00016; EMBL:V00658; NID:G337302

A;Note: this ORF is not annotated in GenBank entry HUMMTTRP, release 106

R;Horai, S.; Hayasaka, K.; Kondo, R.; Tsugane, K.; Takahata, N.

Proc. Natl. Acad. Sci. U.S.A. 92, 532-536, 1995

A;Title: Recent African origin of modern humans revealed by complete sequences of homino-

A;Reference number: I59384; MUID:95132634; PMID:7530363

A;Accession: I80239

A;Status: preliminary; translated from GB/EMBL/DBDJ

A;Molecule type: DNA

A;Residues: 1-210,'V',212-256,'V',258-313,'V',315-455,'R',457-603 <HOR>

A;Cross-references: DDBJ:D38112; NID:g644480; PIDN:BAA07297.1; PID:g704447

A;Experimental source: African isolate SBI7

C;Genetics:

A;Gene: GDB:MTNDS

A;Cross-references: GDB:118916; OMIM:516005

A;Map position: MTH12337-14148

A;Genome: mitochondrion

A;Genetic code: SGC1

C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5

C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 88.2%; Score 112; DB 1; Length 603;
Best Local Similarity 96.0%; Pred. No. 2,9e-08;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLTALDLYPT 26

Db 484 LYLKLTALAVTFGLTALDLYPT 508

NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenas
NADH2 dehydrogenase
NADH2 dehydrogenas

61 48.0 606 2 T11257
60 47.2 605 2 T11203
60 47.2 610 2 T11544
60 47.2 612 2 T14101
60 47.2 612 2 T14104
59 46.5 607 1 OXMS5M
59 46.5 609 2 T11774
58 45.7 602 2 T11438
58 45.7 604 2 CS8889
56 44.1 601 2 T11190
56 44.1 606 2 T45560
56 44.1 612 2 T09867
55 43.3 603 2 T11516
55 43.3 612 2 T09957
54 42.5 606 2 C90627
54 42.5 611 2 CS8893

RESULT 2

T14149
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - pygmy chimpanzee mitochondrion
 C:Species: mitochondrion Pan paniscus (pygmy chimpanzee, bonobo)
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T14149
 R:Horai, S.; Hayasaka, K.; Kondo, R.; Tsugane, K.; Takahata, N.
 Proc. Natl. Acad. Sci. U.S.A. 92, 532-536, 1995
 A:Title: Recent African origin of modern humans revealed by complete sequences of hominoid
 A:Reference number: I59384; MUID:95132634; PMID:7530363
 A:Accession: T14149
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-603 <HOR>
 A:Cross-references: UNIPROT:P03916; EMBL:D38116; PIDN:BAA07315.1
 C:Genetics:
 A:Genome: mitochondrion
 A:Genetic code: SGCI
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 85.8%; Score 109; DB 2; Length 603;
 Best Local Similarity 92.0%; Pred. No. 7.8e-08;
 Matches 23; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 LYLKLTALAVTFGLLTALDNLNPT 26

Db 484 LYLKLTALSVTFGLLTALDNLNLT 508

RESULT 3

B59154
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - western lowland gorilla mitochondrion
 N:Alternate names: NADH-ubiquinone oxidoreductase chain 5
 C:Species: mitochondrion Gorilla gorilla gorilla (western lowland gorilla)
 C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Jun-2002
 C:Accession: B59154
 R:Xu, X.; Arnason, U.
 Mol. Biol. Evol. 13, 691-698, 1996
 A:Title: A complete sequence of the mitochondrial genome of the western lowland gorilla.
 A:Reference number: Z17269; MUID:96212991; PMID:8676744
 A:Accession: B59154
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown; translated
 A:Molecule type: DNA
 A:Residues: 1-603 <XUX>
 A:Cross-references: GB:X93347; NID:q1304307; GSPDB:GN00106
 A:Note: submitted to GenBank, November 1995
 A:Note: this translation is not annotated in GenBank entry GGMITG, release 114.0
 C:Genetics:
 A:Gene: NADH5
 A:Genome: mitochondrion
 A:Genetic code: SGCI
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 80.3%; Score 102; DB 2; Length 603;
 Best Local Similarity 84.0%; Pred. No. 7.9e-07;
 Matches 21; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 LYLKLTALAVTFGLLTALDNLNPT 26

Db 484 LYLKLTALSVTFGLLTALDNLNLT 508

RESULT 4

T14026
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - gorilla mitochondrion
 C:Species: mitochondrion Gorilla gorilla (gorilla)
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
 C:Accession: T14026
 R:Horai, S.; Hayasaka, K.; Kondo, R.; Tsugane, K.; Takahata, N.
 Proc. Natl. Acad. Sci. U.S.A. 92, 532-536, 1995

A:Title: Recent African origin of modern humans revealed by complete sequences of hominoid
 A:Reference number: I59384; MUID:95132634; PMID:7530363
 A:Accession: T14026
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-603 <HOR>
 A:Cross-references: UNIPROT:P03917; EMBL:D38114; PIDN:BAA07306.1
 C:Genetics:
 A:Genome: mitochondrion
 A:Genetic code: SGCI
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 80.3%; Score 102; DB 2; Length 603;
 Best Local Similarity 84.0%; Pred. No. 7.9e-07;
 Matches 21; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 LYLKLTALAVTFGLLTALDNLNPT 26

Db 484 LYLKLTALSVTFGLLTALDNLNLT 508

RESULT 5

T14199
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - chimpanzee mitochondrion
 C:Species: mitochondrion Pan troglodytes (chimpanzee)
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
 C:Accession: T14199
 R:Horai, S.; Hayasaka, K.; Kondo, R.; Tsugane, K.; Takahata, N.
 Proc. Natl. Acad. Sci. U.S.A. 92, 532-536, 1995
 A:Title: Recent African origin of modern humans revealed by complete sequences of hominoid
 A:Reference number: I59384; MUID:95132634; PMID:7530363
 A:Accession: T14199
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-603 <HOR>
 A:Cross-references: UNIPROT:P035648; EMBL:D38113; PIDN:BAA07302.1
 C:Genetics:
 A:Genome: mitochondrion
 A:Genetic code: SGCI
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 78.7%; Score 100; DB 2; Length 603;
 Best Local Similarity 88.0%; Pred. No. 1.5e-06;
 Matches 22; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 2 LYLKLTALAVTFGLLTALDNLNPT 26

Db 484 LYLKLTALSVTFGLLTALDNLNLT 508

RESULT 6

T14145
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - orangutan mitochondrion
 C:Species: mitochondrion Pongo pygmaeus (orangutan)
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
 C:Accession: T14145
 R:Horai, S.; Hayasaka, K.; Kondo, R.; Tsugane, K.; Takahata, N.
 Proc. Natl. Acad. Sci. U.S.A. 92, 532-536, 1995
 A:Title: Recent African origin of modern humans revealed by complete sequences of hominoid
 A:Reference number: I59384; MUID:95132634; PMID:7530363
 A:Accession: T14145
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-603 <HOR>
 A:Cross-references: UNIPROT:P03918; EMBL:D38115; PIDN:BAA07311.1
 C:Genetics:
 A:Genome: mitochondrion
 A:Genetic code: SGCI
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 80.3%; Score 102; DB 2; Length 603;
 Best Local Similarity 84.0%; Pred. No. 7.9e-07;
 Matches 21; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 LYLKLTALAVTFGLLTALDNLNPT 26

Db 484 LYLKLTALSVTFGLLTALDNLNLT 508

Query Match 67.7%; Score 86; DB 2; Length 603;
 Best Local Similarity 78.3%; Pred. No. 0.00015;
 Matches 18; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDLY 24
 ||||| : : : : :
 Db 484 LYLKLAALTATLLGLLVALDLY 506

RESULT 7

T11843
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - common gibbon mitochondrion
 N1Alternate names: NADH-ubiquinone oxidoreductase chain 5
 C:Species: Mitochondrion Hylobates lar (common gibbon, white-handed gibbon)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
 C:Accession: T11843; A00449
 R:Arnason, U.; Gullberg, A.; Xu, X.
 Hereditas 124, 185-189, 1996
 A:Title: A complete mitochondrial DNA molecule of the white-handed gibbon, Hylobates lar
 A:Reference number: Z17353
 A:Accession: T11843
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-603 <ARN>
 A:Cross-references: UNIPROT:P03919; EMBL:X99256; PIDN:CAA67639.1
 A:Experimental source: entry is labelled fragment isolate Ester
 R:Brown, W.M.; Prager, E.M.; Wang, A.; Wilson, A.C.
 J. Mol. Evol. 18, 225-239, 1982
 A:Title: Mitochondrial DNA sequences of primates: tempo and mode of evolution.
 A:Reference number: A00435; MUID:82242101; PMID:6284948
 A:Accession: A00449
 A:Molecule type: DNA
 A:Residues: 1-79 <BRO>
 A:Cross-references: GB:V00659; NID:G12996; PIDN:CAA24025.1; PID:G12997
 C:Genetics:
 A:Gene: NADH5
 A:Genome: mitochondrion
 A:Genetic code: SGC1
 A:Start codon: ATA
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 63.8%; Score 81; DB 2; Length 603;
 Best Local Similarity 68.0%; Pred. No. 0.00079;
 Matches 17; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDLY 26
 ||||| : : : : :
 Db 484 LYLKLTALSITLLGLTAFDLHLT 508

RESULT 8

T11503
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - dog mitochondrion
 C:Species: Mitochondrion Canis lupus familiaris (dog)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 03-Jun-2002
 C:Accession: T11503
 R:Kim, K.S.; Lee, S.E.; Jeong, H.W.; Ha, J.H.
 Mol. Phylogenet. Evol. 10, 210-220, 1998
 A:Title: The complete nucleotide sequence of the domestic dog (Canis familiaris) mitocho
 A:Reference number: Z17276; MUID:99097286; PMID:9878232
 A:Accession: T11503
 A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL
 A:Molecule type: DNA
 A:Residues: 1-606 <KIN>
 A:Cross-references: EMBL:U96639; NID:G4154170; PID:G4154181; PIDN:AAD04773.1
 C:Genetics:
 A:Genome: mitochondrion
 A:Genetic code: SGC1
 A:Note: ND5
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 57.9%; Score 73.5; DB 2; Length 606;
 Best Local Similarity 51.5%; Pred. No. 0.0093;
 Matches 17; Conservative 4; Mismatches 3; Indels 9; Gaps 1;

QY 3 YLKLTAALAVTFGLLTALDLY 26
 ||||| : : : : :
 Db 485 YLKLTAALAVTFGLLTALDLY 517

RESULT 9

S26161
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - harbor seal mitochondrion
 C:Species: Mitochondrion Phoca vitulina (harbor seal)
 C:Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 09-Jul-2004
 C:Accession: S26161
 R:Arnason, U.; Johnsson, E.
 J. Mol. Evol. 34, 493-505, 1992
 A:Title: The complete mitochondrial DNA sequence of the harbor seal, Phoca vitulina.
 A:Reference number: S26151; MUID:92277666; PMID:1593642
 A:Accession: S26161
 A:Molecule type: DNA
 A:Residues: 1-609 <ARN>
 A:Cross-references: UNIPROT:Q00542; EMBL:X63726; NID:G13431; PIDN:CAA45267.1; PID:G13442
 C:Genetics:
 A:Genome: mitochondrion
 A:Genetic code: SGC1
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 57.9%; Score 73.5; DB 2; Length 609;
 Best Local Similarity 51.5%; Pred. No. 0.0094;
 Matches 17; Conservative 3; Mismatches 4; Indels 9; Gaps 1;

QY 3 YLKLTAALAVTFGLLTALDLY 26
 ||||| : : : : :
 Db 488 YLKLTAALAVTFGLLTALDLY 520

RESULT 10

T11412
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - cat mitochondrion
 C:Species: Mitochondrion Felis silvestris catus (domestic cat)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
 C:Accession: T11412
 R:Lopez, J.V.; Cevario, S.; O'Brien, S.J.
 Genomics 33, 229-246, 1996
 A:Title: Complete nucleotide sequences of the domestic cat (Felis catus) mitochondrial g
 A:Reference number: Z17268; MUID:96301400; PMID:8660972
 A:Accession: T11412
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-606 <LOP>
 A:Cross-references: UNIPROT:P48921; EMBL:U20753; NID:G1098523; PID:G1098534; PIDN:AAC485;
 A:Experimental source: female adult; isolate FCA-65; lymphocyte; blood
 C:Genetics:
 A:Genome: mitochondrion
 A:Note: ND5
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 5
 C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Query Match 57.1%; Score 72.5; DB 2; Length 606;
 Best Local Similarity 51.5%; Pred. No. 0.013;
 Matches 17; Conservative 3; Mismatches 4; Indels 9; Gaps 1;

QY 3 YLKLTAALAVTFGLLTALDLY 26
 ||||| : : : : :
 Db 485 YLKLTAALAVTFGLLTALDLY 517

RESULT 11

T10982
 NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - pig mitochondrion
 C:Species: Mitochondrion Sus scrofa domestica (domestic pig)

C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C;Accession: T10982; T11880
R;Lin. C.S.; Liu, C.Y.; Sun, Y.L.; Chang, L.C.; Cheng, I.C.; Yang, P.C.; Mao, S.J.T.; Hu
submitted to the EMBL Data Library, November 1997
A;Description: Complete nucleotide sequence of the porcine mitochondrial genome.
A;Reference number: Z17237
A;Accession: T10982
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-606 <LIN>
A;Cross-references: UNIPROT:Q9TDR1; EMBL:AF034253; NID:g4958951; PID:g4958952; PIDN:AA03
R;Ursing, B.M.
A;Description: The complete mitochondrial DNA sequence of the pig (Sus scrofa).
A;Reference number: Z17370
A;Accession: T11880
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-12, '1', '14-121, 'L', '123-590, 'S', '592-606 <URS>
A;Cross-references: EMBL:AJ002189; PIDN:CAA05241.1
C;Genetics:
A;Gene: NADH5
A;Genome: mitochondrion
A;Genetic code: SGC1
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;
Query Match 56.7%; Score 72; DB 2; Length 606;
Best Local Similarity 56.0%; Pred. No. 0.015;
Matches 14; Conservative 5; Mismatches 6; Indels 0; Gaps 0;
QY 2 LYLKLTALAVTFGLLTALDLYPT 26
DB 484 LYLKLTALAVTFGLLTALDLYPT 508
RESULT 12
T11399
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - fat dormouse mitochondrion
C;Species: mitochondrion Myoxus glis (fat dormouse)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C;Accession: T11399
R;Reyes, A.; Pesole, G.; Saccone, C.
Mol. Biol. Evol. 15, 499-505, 1998
A;Title: Complete mitochondrial DNA sequence of the fat dormouse, Glis glis: further evi
A;Reference number: Z17267; MUID:98242079; PMID:9580978
A;Accession: T11399
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-603 <REY>
A;Cross-references: UNIPROT:O63908; EMBL:AJ001562; NID:g3127895; PIDN:CAA04840.1; PID:g3
A;Gene: NADH5
A;Genome: mitochondrion
A;Genetic code: SGC1
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;
Query Match 55.1%; Score 70; DB 2; Length 603;
Best Local Similarity 66.7%; Pred. No. 0.023;
Matches 14; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 3 YLKLTAALAVTFGLLTALDLYN 23
DB 485 YLKLTAALAVTFGLLTALDLYN 505
RESULT 13
T11060
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - sheep mitochondrion
C;Species: mitochondrion Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C;Accession: T11060

R;Hiendleder, S.; Lewalski, H.; Wassmuth, R.; Janke, A.
J. Mol. Evol. 47, 441-448, 1998
A;Title: The complete mitochondrial DNA sequence of the domestic sheep (Ovis aries) and
A;Reference number: Z17245; MUID:98440761; PMID:9767689
A;Accession: T11060
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-606 <HIE>
A;Cross-references: UNIPROT:O78756; EMBL:AF010406; NID:g3445513; PID:g3366632; PIDN:AA010
A;Experimental source: strain Merinolandschaf; liver
C;Genetics:
A;Genome: mitochondrion
A;Genetic code: SGC1
A;Note: ND5
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;
Query Match 54.7%; Score 69.5; DB 2; Length 606;
Best Local Similarity 45.5%; Pred. No. 0.035;
Matches 15; Conservative 5; Mismatches 4; Indels 9; Gaps 1;
QY 3 YLKLTAALAVTFGLLTALDLYN 26
DB 485 YLKLTAALAVTFGLLTALDLYN 517
RESULT 14
T11451
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - nine-banded armadillo mitochondrion
C;Species: mitochondrion Dasypus novemcinctus (nine-banded armadillo)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C;Accession: T11451
R;Arnason, U.; Gullberg, A.; Janke, A.
Mol. Biol. Evol. 14, 762-768, 1997
A;Title: Phylogenetic analysis of mitochondrial DNA suggest a sister group relationship b
A;Reference number: Z17272; MUID:97357423; PMID:9214749
A;Accession: T11451
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-601 <ARN>
A;Cross-references: UNIPROT:O21335; EMBL:Y11832; NID:g2252500; PIDN:CAA72521.1; PID:g2252
C;Genetics:
A;Gene: NADH5
A;Genome: mitochondrion
A;Genetic code: SGC1
C;Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C;Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;
Query Match 54.3%; Score 69; DB 2; Length 601;
Best Local Similarity 59.1%; Pred. No. 0.041;
Matches 13; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
QY 2 LYLKLTALAVTFGLLTALDLYN 23
DB 483 LYLKLTALAVTFGLLTALDLYN 504
RESULT 15
T11867
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 5 - horse mitochondrion
C;Species: mitochondrion Equus caballus (domestic horse)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C;Accession: T11867
R;Xu, X.; Arnason, U.
Gene 148, 357-362, 1994
A;Title: The complete mitochondrial DNA sequence of the horse, Equus caballus: Extensive
A;Reference number: Z17369; MUID:95047450; PMID:7958969
A;Accession: T11867
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-604 <XUX>
A;Cross-references: UNIPROT:P48656; EMBL:X79547; NID:g577571; PID:g577582; PIDN:CAA56089.
C;Genetics:

A: Genome: mitochondrion
C: Superfamily: NADH dehydrogenase (ubiquinone) chain 5
C: Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;
Query March 54.3%; Score 69; DB 2; Length 604;
Best Local Similarity 62.5%; Pred. NO. 0.041;
Matches 15; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
QY 3 YLKLTALAVTFGLLTALDLYNPT 26
Db 483 YMKLTALAVTILGFTLALSLMT 506

Search completed: November 10, 2004, 13:40:37
Job time : 6.79275 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 31.7478 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-28
Perfect score: 127
Sequence: 1 TLXLKLTALAVTFGLLTALDNLNPT 26

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	112	88.2	601	2 AAT38537	Aat38537 homo sapi
2	112	88.2	601	2 AAT38538	Aat38538 homo sapi
3	112	88.2	603	1 NU5M HUMAN	P03915 homo sapien
4	112	88.2	603	2 QPZ35	Q6p235 homo sapien
5	112	88.2	603	2 QSRM4	Q6rmf4 homo sapien
6	112	88.2	603	2 QSRM4	Q6rmf4 homo sapien
7	112	88.2	603	2 QSRM5	Q6rmf5 homo sapien
8	112	88.2	603	2 QSRN10	Q6rnj0 homo sapien
9	112	88.2	603	2 QSRN11	Q6rnj1 homo sapien
10	112	88.2	603	2 QSRPF2	Q6rpf2 homo sapien
11	112	88.2	603	2 QSRQ14	Q6rq14 homo sapien
12	112	88.2	603	2 QSRR80	Q6rrt0 homo sapien
13	112	88.2	603	2 QSR800	Q6rs00 homo sapien
14	112	88.2	603	2 QSR527	Q6rs27 homo sapien
15	112	88.2	603	2 QSVHD5	Q6vhd5 homo sapien
16	112	88.2	603	2 QSVI79	Q6vi79 homo sapien
17	112	88.2	603	2 QSVI05	Q6vid5 homo sapien
18	112	88.2	603	2 QSVIF3	Q6vif3 homo sapien
19	112	88.2	603	2 QSVIX6	Q6vix6 homo sapien
20	112	88.2	603	2 QSVIZ4	Q6viz4 homo sapien
21	112	88.2	603	2 QSVKL7	Q6vkl7 homo sapien
22	112	88.2	603	2 QSQW97	Q6wq97 homo sapien
23	112	88.2	603	2 Q7Y6H2	Q7y6h2 homo sapien
24	112	88.2	603	2 Q7Y622	Q7y622 homo sapien
25	112	88.2	603	2 Q7Y7F0	Q7y7f0 homo sapien
26	112	88.2	603	2 Q7Y823	Q7y823 homo sapien
27	112	88.2	603	2 Q7YCC4	Q7yc4 homo sapien
28	112	88.2	603	2 Q7YCD7	Q7ycd7 homo sapien
29	112	88.2	603	2 Q7YCD9	Q7ycd9 homo sapien
30	112	88.2	603	2 Q7YCS8	Q7yces homo sapien
31	112	88.2	603	2 Q7YCF6	Q7ycf6 homo sapien

ALIGNMENTS

RESULT 1

AAT38537 PRELIMINARY; PRT; 601 AA.

AC AAT38537; (TREMELrel. 27, Created)
DT 01-JUN-2004 (TREMELrel. 27, Last sequence update)
DT 01-JUN-2004 (TREMELrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 5.
GN ND5.

OS Homo sapiens (Human).

OG Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GD7809;

RX MEDLINE=22807936; PubMed=12870132;

RA Kong Q.-P., Yao Y.-G., Sun C., Bandelt H.-J., Zhu C.-L., Zhang Y.-P.;

RT "Phylogeny of east Asian mitochondrial DNA lineages inferred from

RT complete sequences."

RL At. J. Hum. Genet. 73:671-676(2003).

DR ENBL; AY255168; AAT38537.1; -.

RW Mitochondrion.

SQ SEQUENCE 601 AA; 66837 MW; 6CA20C0C96FBD3B4 CRC64;

Query Match 88.2%; Score 112; DB 2; Length 601;

Best Local Similarity 96.0%; Pred. No. 2.8e-07;

Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LXLKLTALAVTFGLLTALDNLNPT 26

Db 482 LXLKLTALAVTFGLLTALDNLNLT 506

RESULT 2

AAT38538 PRELIMINARY; PRT; 601 AA.

AC AAT38538; (TREMELrel. 27, Created)
DT 01-JUN-2004 (TREMELrel. 27, Last sequence update)
DT 01-JUN-2004 (TREMELrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 5.
GN ND5.

OS Homo sapiens (Human).

OG Mitochondrion.

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=QD8147;

RX MEDLINE=22807936; PubMed=12870132;

RA Kong Q.-P., Yao Y.-G., Sun C., Bandelt H.-J., Zhu C.-L., Zhang Y.-P.;

Ingram M., Gyllenstein U.;
"Mitochondrial genome variation and evolutionary history of Australian
and new guinean aborigines.";
Genome Res. 13:1600-1606(2003).
[8]
SEQUENCE FROM N.A.
RP PubMed=14760490; DOI=10.1007/s00414-004-0427-6;
Cobie M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
Irwin J.A., Parsons T.J.;
"Single nucleotide polymorphisms over the entire mtDNA genome that
increase the power of forensic testing in Caucasians.";
Int. J. Legal Med. 118:137-146(2004).
[9]
SEQUENCE OF 1-79 FROM N.A.
RP MEDLINE=82242101; PubMed=6284948;
Brown W.M., Prager E.M., Wang A., Wilson A.C.;
"Mitochondrial DNA sequences of primates: tempo and mode of
evolution.";
J. Mol. Evol. 18:225-239(1982).
[10]
IDENTIFICATION OF PROTEIN.
RP MEDLINE=85188293; PubMed=3921850;
Chomyn A., Mariotti P., Cleeter M.W.J., Ragen C.I., Matsuno-Yagi A.,
Hatefi Y., Doolittle R.F., Attardi G.;
"Six unidentified reading frames of human mitochondrial DNA encode
components of the respiratory-chain NADH dehydrogenase.";
Nature 314:592-597(1985).
[11]
VARIANTS SER-17; SER-95; PRO-99; ASP-146; VAL-160; SER-165; SER-304;
ALA-331 AND GLY-503.
RP MEDLINE=92098084; PubMed=1757091;
Marruki S., Noer A.S., Leiritz P., Thyagarajan D., Kapsa R.,
Utthanaphol P., Byrne B.;
"Normal variants of human mitochondrial DNA and translation products:
the building of a reference data base.";
Hum. Genet. 88:139-145(1991).
[12]
VARIANT LHON THR-458.
RP MEDLINE=92120513; PubMed=1732158;
Brown M.D., Voljavec A.S., Lott M.T., Torroni A., Yang C.C.,
Wallace D.C.;
"Mitochondrial DNA complex I and III mutations associated with Leber's
hereditary optic neuropathy.";
Genetics 130:163-173(1992).
[13]
VARIANT LHON THR-458.
RP MEDLINE=91144615; PubMed=1900003;
Johns D.R., Berman J.;
"Alternative simultaneous complex I mitochondrial DNA mutations in
Leber's hereditary optic neuropathy.";
Biochem. Biophys. Res. Commun. 174:1324-1330(1991).
[14]
VARIANT LHON GLU-465.
RP MEDLINE=94027064; PubMed=8213825;
Howell N., Halvorson S., Burns J., McCullough D.A., Poulton J.;
"When does bilateral optic atrophy become Leber hereditary optic
neuropathy?";
Am. J. Hum. Genet. 53:959-963(1993).
[15]
VARIANT THR-475.
RP MEDLINE=98127994; PubMed=9461455;
Rieder M.J., Taylor S.L., Toke S.L., Nickerson D.A.;
"Automating the identification of DNA variations using quality-based
fluorescence re-sequencing: analysis of the human mitochondrial
genome.";
Nucleic Acids Res. 26:967-973(1998).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein, mitochondrial
inner membrane (probable).
CC -!- DISEASE: Defects in MTND5 are a cause of Leber's hereditary optic
neuropathy (LHON) [MIM:535000]; also known as Leber's optic
atrophy, LHON is a maternally inherited disease resulting in acute
bilateral blindness due to retinal degeneration predominantly in
CC CC

CC young men. Cardiac conduction defects and neurological defects
CC have also been described, resulting in optic nerve degeneration
CC and cardiac dysrhythmia.
CC -!- SIMILARITY: Belongs to the complex I subunit 5 family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J01415; AAB58953.1; -
CC EMBL; V06662; CRA24036.1; -
CC EMBL; D38112; BAA07297.1; -
CC EMBL; AF465942; AANI4556.1; -
CC EMBL; AY339402; AAP89046.1; -
CC EMBL; AY339403; AAP89059.1; -
CC EMBL; AY339404; AAP89072.1; -
CC EMBL; AY339405; AAP89085.1; -
CC EMBL; AY339406; AAP89098.1; -
CC EMBL; AY339407; AAP89111.1; -
CC EMBL; AY339408; AAP89124.1; -
CC EMBL; AY339409; AAP89137.1; -
CC EMBL; AY339410; AAP89150.1; -
CC EMBL; AY339411; AAP89163.1; -
CC EMBL; AY339412; AAP89176.1; -
CC EMBL; AY339413; AAP89189.1; -
CC EMBL; AY339414; AAP89202.1; -
CC EMBL; AY339415; AAP89215.1; -
CC EMBL; AY339416; AAP89228.1; -
CC EMBL; AY339417; AAP89241.1; -
CC EMBL; AY339418; AAP89254.1; -
CC EMBL; AY339419; AAP89267.1; -
CC EMBL; AY339420; AAP89280.1; -
CC EMBL; AY339421; AAP89293.1; -
CC EMBL; AY339422; AAP89306.1; -
CC EMBL; AY339423; AAP89319.1; -
CC EMBL; AY339424; AAP89332.1; -
CC EMBL; AY339425; AAP89345.1; -
CC EMBL; AY339426; AAP89358.1; -
CC EMBL; AY339427; AAP89371.1; -
CC EMBL; AY339428; AAP89384.1; -
CC EMBL; AY339429; AAP89397.1; -
CC EMBL; AY339430; AAP89410.1; -
CC EMBL; AY339431; AAP89423.1; -
CC EMBL; AY339432; AAP89436.1; -
CC EMBL; AY339433; AAP89449.1; -
CC EMBL; AY339434; AAP89462.1; -
CC EMBL; AY339435; AAP89475.1; -
CC EMBL; AY339436; AAP89488.1; -
CC EMBL; AY339437; AAP89501.1; -
CC EMBL; AY339438; AAP89514.1; -
CC EMBL; AY339439; AAP89527.1; -
CC EMBL; AY339440; AAP89540.1; -
CC EMBL; AY339441; AAP89553.1; -
CC EMBL; AY339442; AAP89566.1; -
CC EMBL; AY339443; AAP89579.1; -
CC EMBL; AY339444; AAP89592.1; -
CC EMBL; AY339445; AAP89605.1; -
CC EMBL; AY339446; AAP89618.1; -
CC EMBL; AY339447; AAP89631.1; -
CC EMBL; AY339448; AAP89644.1; -
CC EMBL; AY339449; AAP89657.1; -
CC EMBL; AY339451; AAP89683.1; -
CC EMBL; AY339452; AAP89696.1; -
CC EMBL; AY339453; AAP89709.1; -
CC EMBL; AY339454; AAP89722.1; -
CC EMBL; AY339455; AAP89735.1; -
CC EMBL; AY339456; AAP89748.1; -
CC EMBL; AY339457; AAP89761.1; -

Query Match 88.2%; Score 112; DB 1; Length 603;
Best Local Similarity 96.0%; Pred. No. 2.9e-07;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 2 LYLKLTALAVTFGLLTALDNLNPT 26
Db 484 LYLKLTALAVTFGLLTALDNLNPT 508

RESULT 4
Q6PZ35 PRELIMINARY; PRT; 603 AA.
ID Q6PZ35
AC Q6PZ35; 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 5.
GN Name=NDS;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Starikovskaya E.B., Sukernik R.I., Derbeneva O.A., Volodko N.V.,
RA Torroni A., Ruiz-Pesini E., Brown M.D., Lott M.T., Hosseini S.H.,
RA Huoponen K., Wallace D.C.;
RT "Mitochondrial DNA Diversity in Indigenous Populations of Southern
RT Extent of Siberia, and the Origins of Native American Haplogroups.";
RL Ann. Hum. Genet. 0:0-0(2004).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein, mitochondrial
CC inner membrane (By similarity).
CC -!- SIMILARITY: Belongs to the complex I subunit 5 family.
DR EMBL; AY570524; AAS83326.1; -
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR010934; NADH5_C.
DR InterPro; IPR003945; NADH5_C.
DR InterPro; IPR003916; NADH5_C.
DR InterPro; IPR001750; Oxidored_g1.
DR InterPro; IPR001516; Oxidored_g1.
DR Pfam; PF06455; NADH5_C; 1.
DR Pfam; PF00361; Oxidored_g1; 1.
DR Pfam; PF00662; Oxidored_g1; 1.
DR PRINTS; PR01434; NADH5_C.
DR PRINTS; PR01435; NADH5_C.
DR TIGRFS; TIGR01974; NDH_I_L; 1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 603 AA; 67030 MW; 1DAD00F47AD7B89A CRC64;

Query Match 88.2%; Score 112; DB 2; Length 603;
Best Local Similarity 96.0%; Pred. No. 2.9e-07;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 2 LYLKLTALAVTFGLLTALDNLNPT 26
Db 484 LYLKLTALAVTFGLLTALDNLNPT 508

RESULT 5
Q6RMP4 PRELIMINARY; PRT; 603 AA.
ID Q6RMP4
AC Q6RMP4; 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 5.
GN Name=NDS;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
PubMed=14760490;
Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
Irwin J.A., Parsons T.J.;
"Single nucleotide polymorphisms over the entire mtDNA genome that
increase the power of forensic testing in Caucasians.";
Int. J. Legal Med. 0:0-0(2004).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein, mitochondrial
inner membrane (By similarity).
CC -!- SIMILARITY: Belongs to the complex I subunit 5 family.
DR EMBL; AY495265; AAR94781.1; -;
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR010934; NADH5_C.
DR InterPro; IPR003945; NADHpl_oxred5.
DR InterPro; IPR003916; NADHub_oxred5.
DR InterPro; IPR001750; Oxidored_g1_N.
DR InterPro; IPR001516; Oxidored_g1_N.
DR Pfam; PF06455; NADH5_C; 1.
DR Pfam; PF00361; Oxidored_g1_N; 1.
DR PRINTS; PR01434; NADHDHGNASE5.
DR PRINTS; PR01435; NPOXDRDTASE5.
DR TIGRfams; TIGR01974; NDH_I_1; 1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 603 AA; 67010 MW; 7561BE3A77C9023F CRC64;

Query Match 88.2%; Score 112; DB 2; Length 603;
Best Local Similarity 96.0%; Pred. No. 2.9e-07;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDNLNPT 26
DB 484 LYLKLTALAVTFGLLTALDNLNPT 508

RESULT 6
Q6RMK4 PRELIMINARY; PRT; 603 AA.
AC Q6RMK4;
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DE NADH dehydrogenase subunit 5.
GN Name=NDS;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
Irwin J.A., Parsons T.J.;
"Single nucleotide polymorphisms over the entire mtDNA genome that
increase the power of forensic testing in Caucasians.";
Int. J. Legal Med. 0:0-0(2004).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein, mitochondrial
inner membrane (By similarity).
CC -!- SIMILARITY: Belongs to the complex I subunit 5 family.
DR EMBL; AY495265; AAR94781.1; -;
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR010934; NADH5_C.
DR InterPro; IPR003945; NADHpl_oxred5.
DR InterPro; IPR003916; NADHub_oxred5.
DR InterPro; IPR001750; Oxidored_g1_N.
DR InterPro; IPR001516; Oxidored_g1_N.
DR Pfam; PF06455; NADH5_C; 1.
DR Pfam; PF00361; Oxidored_g1_N; 1.
DR PRINTS; PR01434; NADHDHGNASE5.
DR PRINTS; PR01435; NPOXDRDTASE5.
DR TIGRfams; TIGR01974; NDH_I_1; 1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 603 AA; 67010 MW; 7561BE3A77C9023F CRC64;

Query Match 88.2%; Score 112; DB 2; Length 603;
Best Local Similarity 96.0%; Pred. No. 2.9e-07;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDNLNPT 26
DB 484 LYLKLTALAVTFGLLTALDNLNPT 508

RESULT 7
Q6RMK5 PRELIMINARY; PRT; 603 AA.
AC Q6RMK5;
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DE NADH dehydrogenase subunit 5.
GN Name=NDS;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
Irwin J.A., Parsons T.J.;
"Single nucleotide polymorphisms over the entire mtDNA genome that
increase the power of forensic testing in Caucasians.";
Int. J. Legal Med. 0:0-0(2004).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein, mitochondrial
inner membrane (By similarity).
CC -!- SIMILARITY: Belongs to the complex I subunit 5 family.
DR EMBL; AY495265; AAR94781.1; -;
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR010934; NADH5_C.
DR InterPro; IPR003945; NADHpl_oxred5.
DR InterPro; IPR003916; NADHub_oxred5.
DR InterPro; IPR001750; Oxidored_g1_N.
DR InterPro; IPR001516; Oxidored_g1_N.
DR Pfam; PF06455; NADH5_C; 1.
DR Pfam; PF00361; Oxidored_g1_N; 1.
DR PRINTS; PR01434; NADHDHGNASE5.
DR PRINTS; PR01435; NPOXDRDTASE5.
DR TIGRfams; TIGR01974; NDH_I_1; 1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 603 AA; 67012 MW; 567A22F7E813C766 CRC64;

Query Match 88.2%; Score 112; DB 2; Length 603;
Best Local Similarity 96.0%; Pred. No. 2.9e-07;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDNLNPT 26
DB 484 LYLKLTALAVTFGLLTALDNLNPT 508

RESULT 8
Q6RMJ0 PRELIMINARY; PRT; 603 AA.
AC Q6RMJ0;
ID Q6RMJ0
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 603 AA; 67012 MW; 567A22F7E813C766 CRC64;

Query Match 88.2%; Score 112; DB 2; Length 603;
Best Local Similarity 96.0%; Pred. No. 2.9e-07;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDNLNPT 26
DB 484 LYLKLTALAVTFGLLTALDNLNPT 508

RESULT 9
Q6RMJ1 PRELIMINARY; PRT; 603 AA.
AC Q6RMJ1;
ID Q6RMJ1
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 603 AA; 67012 MW; 567A22F7E813C766 CRC64;

Pfam; PF00662; Oxidored_g1_N; 1.
DR PRINTS; PR01434; NADHDHGNASE5.
DR PRINTS; PR01435; NPOXDRDTASE5.
DR TIGRfams; TIGR01974; NDH_I_1; 1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 603 AA; 67014 MW; 18A0623B7A04162A CRC64;

Query Match 88.2%; Score 112; DB 2; Length 603;
Best Local Similarity 96.0%; Pred. No. 2.9e-07;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDNLNPT 26
DB 484 LYLKLTALAVTFGLLTALDNLNPT 508

RESULT 7
Q6RMK5 PRELIMINARY; PRT; 603 AA.
AC Q6RMK5;
DT 05-JUL-2004 (TREMELrel. 27, Created)
DT 05-JUL-2004 (TREMELrel. 27, Last sequence update)
DE NADH dehydrogenase subunit 5.
GN Name=NDS;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
Irwin J.A., Parsons T.J.;
"Single nucleotide polymorphisms over the entire mtDNA genome that
increase the power of forensic testing in Caucasians.";
Int. J. Legal Med. 0:0-0(2004).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein, mitochondrial
inner membrane (By similarity).
CC -!- SIMILARITY: Belongs to the complex I subunit 5 family.
DR EMBL; AY495265; AAR94781.1; -;
DR EMBL; AY495266; AAR94782.1; -;
DR EMBL; AY495267; AAR94783.1; -;
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR010934; NADH5_C.
DR InterPro; IPR003945; NADHpl_oxred5.
DR InterPro; IPR003916; NADHub_oxred5.
DR InterPro; IPR001750; Oxidored_g1_N.
DR InterPro; IPR001516; Oxidored_g1_N.
DR Pfam; PF06455; NADH5_C; 1.
DR Pfam; PF00361; Oxidored_g1_N; 1.
DR PRINTS; PR01434; NADHDHGNASE5.
DR PRINTS; PR01435; NPOXDRDTASE5.
DR TIGRfams; TIGR01974; NDH_I_1; 1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 603 AA; 67012 MW; 567A22F7E813C766 CRC64;

Query Match 88.2%; Score 112; DB 2; Length 603;
Best Local Similarity 96.0%; Pred. No. 2.9e-07;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDNLNPT 26
DB 484 LYLKLTALAVTFGLLTALDNLNPT 508

RESULT 8
Q6RMJ0 PRELIMINARY; PRT; 603 AA.
AC Q6RMJ0;
ID Q6RMJ0
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 603 AA; 67012 MW; 567A22F7E813C766 CRC64;


```
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 5.
GN Name=ND5;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
RA Irwin J.A., Parsons T.J.;
RT "Single nucleotide polymorphisms over the entire mtDNA genome that
RT increase the power of forensic testing in Caucasians.";
RL Int. J. Legal Med. 0:0-0(2004).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein, mitochondrial
CC inner membrane (By similarity).
CC -!- SIMILARITY: Belongs to the complex I subunit 5 family.
DR EMBL; AY495226; AAR94274.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR010934; NADH5_C.
DR InterPro; IPR003945; NADHpl_oxred5.
DR InterPro; IPR003916; NADHpl_oxred5.
DR Name=ND5;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
RA Irwin J.A., Parsons T.J.;
RT "Single nucleotide polymorphisms over the entire mtDNA genome that
RT increase the power of forensic testing in Caucasians.";
RL Int. J. Legal Med. 0:0-0(2004).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein, mitochondrial
CC inner membrane (By similarity).
CC -!- SIMILARITY: Belongs to the complex I subunit 5 family.
DR EMBL; AY495226; AAR94274.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR010934; NADH5_C.
DR InterPro; IPR003945; NADHpl_oxred5.
DR InterPro; IPR003916; NADHpl_oxred5.
DR InterPro; IPR001750; Oxidored_g1.
DR Pfam; PF00662; Oxidored_g1_N; 1.
DR Pfam; PF00361; Oxidored_g1_N; 1.
DR PRINTS; PR01434; NADHpl_oxred5.
DR PRINTS; PR01435; NPOXDRDTASES.
DR TIGRFAMs; TIGR01974; NDH I_L; 1.
DR Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 603 AA; 67068 MW; 19EC35C76A491631 CRC64;

Query Match 88.2%; Score 112; DB 2; Length 603;
Best Local Similarity 96.0%; Pred. No. 2.9e-07;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDNLNPT 26
DB 484 LYLKLTALAVTFGLLTALDNLNLT 508

RESULT 9
ID Q6RNL1 PRELIMINARY; PRT; 603 AA.
AC Q6RNL1;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 5.
GN Name=ND5;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
RA Irwin J.A., Parsons T.J.;
RT "Single nucleotide polymorphisms over the entire mtDNA genome that
RT increase the power of forensic testing in Caucasians.";
RL Int. J. Legal Med. 0:0-0(2004).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein, mitochondrial
CC inner membrane (By similarity).
CC -!- SIMILARITY: Belongs to the complex I subunit 5 family.
DR EMBL; AY495195; AAR93871.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR010934; NADH5_C.
DR InterPro; IPR003945; NADHpl_oxred5.
DR InterPro; IPR003916; NADHpl_oxred5.
DR InterPro; IPR001750; Oxidored_g1.
DR InterPro; IPR001516; Oxidored_g1_N.
DR Pfam; PF06455; NADH5_C; 1.
DR Pfam; PF00361; Oxidored_g1_N; 1.
DR PRINTS; PR01434; NADHpl_oxred5.
DR PRINTS; PR01435; NPOXDRDTASES.
DR TIGRFAMs; TIGR01974; NDH I_L; 1.
DR Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 603 AA; 67068 MW; 19EC35C76A491631 CRC64;

Query Match 88.2%; Score 112; DB 2; Length 603;
Best Local Similarity 96.0%; Pred. No. 2.9e-07;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDNLNPT 26
DB 484 LYLKLTALAVTFGLLTALDNLNLT 508

RESULT 10
ID Q6RPF2 PRELIMINARY; PRT; 603 AA.
AC Q6RPF2;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE NADH dehydrogenase subunit 5.
GN Name=ND5;
OS Homo sapiens (Human).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14760490;
RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
RA Irwin J.A., Parsons T.J.;
RT "Single nucleotide polymorphisms over the entire mtDNA genome that
RT increase the power of forensic testing in Caucasians.";
RL Int. J. Legal Med. 0:0-0(2004).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein, mitochondrial
CC inner membrane (By similarity).
CC -!- SIMILARITY: Belongs to the complex I subunit 5 family.
DR EMBL; AY495195; AAR93871.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR InterPro; IPR010934; NADH5_C.
DR InterPro; IPR003945; NADHpl_oxred5.
DR InterPro; IPR003916; NADHpl_oxred5.
DR InterPro; IPR001750; Oxidored_g1.
DR InterPro; IPR001516; Oxidored_g1_N.
DR Pfam; PF06455; NADH5_C; 1.
DR Pfam; PF00361; Oxidored_g1_N; 1.
DR PRINTS; PR01434; NADHpl_oxred5.
DR PRINTS; PR01435; NPOXDRDTASES.
DR TIGRFAMs; TIGR01974; NDH I_L; 1.
DR Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
SQ SEQUENCE 603 AA; 67113 MW; 8F3D18CCA57D09DE CRC64;

Query Match 88.2%; Score 112; DB 2; Length 603;
Best Local Similarity 96.0%; Pred. No. 2.9e-07;
Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDNLNPT 26
DB 484 LYLKLTALAVTFGLLTALDNLNLT 508
```


Query Match 88.2%; Score 112; DB 2; Length 603;
 Best Local Similarity 96.0%; Pred. No. 2.9e-07;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDLYNPT 26
 |||||
 Db 484 LYLKLTALAVTFGLLTALDLYNLT 508

RESULT 14
 Q6RS27 PRELIMINARY; PRT; 603 AA.
 AC Q6RS27;
 DT 05-JUL-2004 (TRENBLrel. 27, Created)
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE NADH dehydrogenase subunit 5.
 GN Name=ND5;
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=14760490;
 RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
 RA Irwin J.A., Parsons T.J.;
 RA "Single nucleotide polymorphisms over the entire mtDNA genome that
 RT increase the power of forensic testing in Caucasians.";
 RL Int. J. Legal Med. 0:0-0(2004).
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein, mitochondrial
 CC inner membrane (By similarity).
 CC -!- SIMILARITY: Belongs to the complex I subunit 5 family.
 DR EMBL; AY495093; AAR92545.1; -;
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR InterPro; IPR010934; NADH5_C.
 DR InterPro; IPR003945; NADHpl_oxred5.
 DR InterPro; IPR003916; NADHpl_oxred5.
 DR InterPro; IPR001750; Oxidored_q1.
 DR InterPro; IPR001516; Oxidored_q1_N.
 DR Pfam; PF06455; NADH5_C; 1.
 DR Pfam; PF00361; Oxidored_q1; 1.
 DR Pfam; PF00662; Oxidored_q1_N; 1.
 DR PRINTS; PR01434; NADHGNASES.
 DR PRINTS; PR01435; NPOXDRDTASES.
 DR TIGRfams; TIGR01974; NDH_I_L; 1.
 KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
 SQ SEQUENCE 603 AA; 67038 MW; 032A41FA9D6E18DA CRC64;

Query Match 88.2%; Score 112; DB 2; Length 603;
 Best Local Similarity 96.0%; Pred. No. 2.9e-07;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDLYNPT 26
 |||||
 Db 484 LYLKLTALAVTFGLLTALDLYNLT 508

RESULT 15
 Q6VHD5 PRELIMINARY; PRT; 603 AA.
 AC Q6VHD5;
 DT 05-JUL-2004 (TRENBLrel. 27, Created)
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
 DE NADH dehydrogenase subunit 5.
 GN Name=ND5;
 OS Homo sapiens (Human).
 OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21313109; PubMed=11349229;
 RA Finnila S., Lehtonen M.S., Majamaa K.;
 RA "Phylogenetic network for European mtDNA.";
 RL Am. J. Hum. Genet. 68:1475-1484(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX PubMed=12949126;
 RA Moilanen J.S., Finnila S., Majamaa K.;
 RA "Lineage-specific selection in human mtDNA: lack of polymorphisms in a
 RT segment of MTND5 gene in haplogroup J.";
 RL Mol. Biol. Evol. 20:2132-2142(2003).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX PubMed=14760490;
 RA Coble M.D., Just R.S., O'Callaghan J.E., Letmanyi I.H., Peterson C.T.,
 RA Irwin J.A., Parsons T.J.;
 RA "Single nucleotide polymorphisms over the entire mtDNA genome that
 RT increase the power of forensic testing in Caucasians.";
 RL Int. J. Legal Med. 0:0-0(2004).
 CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein, mitochondrial
 CC inner membrane (By similarity).
 CC -!- SIMILARITY: Belongs to the complex I subunit 5 family.
 DR EMBL; AY339582; AAP91386.1; -;
 DR EMBL; AY495231; AAR94339.1; -;
 DR EMBL; AY495232; AAR94352.1; -;
 DR EMBL; AY495233; AAR94365.1; -;
 DR EMBL; AY495234; AAR94378.1; -;
 DR EMBL; AY495235; AAR94391.1; -;
 DR EMBL; AY495236; AAR94404.1; -;
 DR EMBL; AY495237; AAR94417.1; -;
 DR EMBL; AY495238; AAR94430.1; -;
 DR EMBL; AY339581; AAP91373.1; -;
 DR GO; GO:0005739; C:mitochondrion; IEA.
 DR InterPro; IPR010934; NADH5_C.
 DR InterPro; IPR003945; NADHpl_oxred5.
 DR InterPro; IPR003916; NADHpl_oxred5.
 DR InterPro; IPR001750; Oxidored_q1.
 DR InterPro; IPR001516; Oxidored_q1_N.
 DR Pfam; PF06455; NADH5_C; 1.
 DR Pfam; PF00361; Oxidored_q1; 1.
 DR Pfam; PF00662; Oxidored_q1_N; 1.
 DR PRINTS; PR01434; NADHGNASES.
 DR PRINTS; PR01435; NPOXDRDTASES.
 DR TIGRfams; TIGR01974; NDH_I_L; 1.
 KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
 SQ SEQUENCE 603 AA; 67066 MW; 1AD931F9997A2531 CRC64;

Query Match 88.2%; Score 112; DB 2; Length 603;
 Best Local Similarity 96.0%; Pred. No. 2.9e-07;
 Matches 24; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LYLKLTALAVTFGLLTALDLYNPT 26
 |||||
 Db 484 LYLKLTALAVTFGLLTALDLYNLT 508

Search completed: November 10, 2004, 13:38:24
 Job time : 32.7478 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:32:37 ; Search time 15.1986 Seconds
(without alignments)
191.991 Million cell updates/sec

Title: US-10-092-750-29

Perfect score: 229

Sequence: 1 AGVSAEPSPFPQTRRSWVF.....REVGEPRSRHLNSTDDAE 44

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/6C_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/6D_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	58.5	25.5	396	4	US-09-107-532A-5033
2	57.5	25.1	687	4	US-09-252-991A-3150
3	56.5	24.7	376	4	US-09-107-532A-4633
4	56.5	24.7	506	4	US-09-252-991A-32150
5	55.5	24.2	407	4	US-09-252-991A-21346
6	55.5	24.2	409	4	US-09-252-991A-21338
7	54.5	23.8	373	4	US-09-134-000C-4666
8	54.5	23.8	373	4	US-09-540-236-1942
9	54.5	23.8	408	3	US-09-140-466-4
10	54	23.6	242	1	US-08-599-480-4
11	54	23.6	267	4	US-09-252-991A-32486
12	54	23.6	339	4	US-09-187-999-15
13	54	23.6	474	4	US-09-248-796A-25524
14	53.5	23.4	409	3	US-09-140-466-5
15	53.5	23.4	433	4	US-09-252-991A-32205
16	53.5	23.4	502	1	US-08-496-855A-4
17	53.5	23.4	502	2	US-08-466-589-10
18	53.5	23.4	502	3	US-08-700-636-10
19	53.5	23.4	502	3	US-08-467-574-10
20	53.5	23.4	502	3	US-09-217-345-10
21	53.5	23.4	502	4	US-08-487-596-14
22	53.5	23.4	502	4	US-08-660-451A-14
23	53.5	23.4	502	4	US-08-892-985-10
24	52	22.7	514	3	US-09-066-047-3
25	52	22.7	611	4	US-09-252-991A-17597
26	51.5	22.5	146	3	US-08-936-165A-349
27	51.5	22.5	229	4	US-09-543-681A-6896

28	51.5	22.5	310	4	US-09-710-279-1312	Sequence 1312, Ap
29	51.5	22.5	370	4	US-09-489-039A-11795	Sequence 11795, A
30	51.5	22.5	394	3	US-09-140-466-6	Sequence 6, Appl1
31	51.5	22.5	394	4	US-09-218-197-2	Sequence 2, Appl1
32	51.5	22.5	394	4	US-09-492-709A-247	Sequence 247, App
33	51.5	22.5	414	3	US-09-134-001C-5377	Sequence 5377, Ap
34	51.5	22.5	417	4	US-09-489-039A-10819	Sequence 10819, A
35	51.5	22.5	1028	4	US-09-543-681A-7181	Sequence 7181, Ap
36	51.5	22.5	2532	3	US-09-215-694-10	Sequence 10, Appl
37	51	22.3	27	2	US-08-706-741B-88	Sequence 88, Appl
38	51	22.3	27	2	US-08-924-695A-88	Sequence 88, Appl
39	51	22.3	214	3	US-09-208-718-1	Sequence 1, Appl1
40	51	22.3	664	4	US-09-252-991A-31350	Sequence 31350, A
41	51	22.3	1275	3	US-09-120-513-2	Sequence 2, Appl1
42	51	22.3	1275	3	US-09-450-105-2	Sequence 2, Appl1
43	50.5	22.1	384	4	US-09-328-352-5673	Sequence 5673, Ap
44	50	21.8	59	2	US-08-733-505A-55	Sequence 55, Appl
45	50	21.8	59	2	US-08-733-505A-56	Sequence 56, Appl

ALIGNMENTS

RESULT 1

US-09-107-532A-5033
Sequence 5033 Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD/ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Atinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5033:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...396
SEQUENCE DESCRIPTION: SEQ ID NO: 5033:
US-09-107-532A-5033

```

; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (781)893-5007
;   TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4633:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 376 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (B) LOCATION 1...376
; SEQUENCE DESCRIPTION: SEQ ID NO: 4633:
US-09-107-532A-4633

Query Match      24.7%; Score 56.5; DB 4; Length 376;
Best Local Similarity 35.7%; Pred.No.9.2; 18; Indels 3; Gaps 1;
Matches 15; Conservative 6; Mismatches 10; Indels 11; Gaps 1;

QY    3 VFSAPSPFPQTRSRMVFAHRLRYGVDFRSHLNSTDDAE 44
      |||||:::||||:||||:||||:||||:||||:
Db     86 VVSAADGMPQTFREHILLS--RQGVGPYIVVFLENKVDWDD 124

RESULT 4
US-09-252-991A-32150
; Sequence 32150, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32150
; LENGTH: 506
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32150

Query Match      24.7%; Score 56.5; DB 4; Length 506;
Best Local Similarity 33.3%; Pred.No.13;
Matches 13; Conservative 5; Mismatches 10; Indels 11; Gaps 1;

QY    8 PSFPQTRSRMVFAHRLREV-----GDFFRSRH 35
      |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db     164 PLGVFETAALVFLRRHQGRDPRRTGGGQCHRAAH 202

RESULT 5
US-09-252-991A-21346
; Sequence 21346, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

```

; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21346
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21346
Query Match 24.2%; Score 55.5; DB 4; Length 407;
Best Local Similarity 35.7%; Pred. No. 14;
Matches 15; Conservative 6; Mismatches 18; Indels 3; Gaps 1;
QY 3 VFSAEPSPPQTRSRMVFAHRLREVGDEFRSHLNSTDDADE 44
DB 115 VCSAAGPMPQTRHILLS---RQGVPIVVFVFLNKADMVDD 153
RESULT 6
US-09-252-991A-21338
; Sequence 21338, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21338
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21338
Query Match 24.2%; Score 55.5; DB 4; Length 409;
Best Local Similarity 35.7%; Pred. No. 14;
Matches 15; Conservative 6; Mismatches 18; Indels 3; Gaps 1;
QY 3 VFSAEPSPPQTRSRMVFAHRLREVGDEFRSHLNSTDDADE 44
DB 117 VCSAAGPMPQTRHILLS---RQGVPIVVFVFLNKADMVDD 155
RESULT 7
US-09-134-000C-4666
; Sequence 4666, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 4666
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4666
Query Match 23.8%; Score 54.5; DB 4; Length 373;
Best Local Similarity 35.7%; Pred. No. 17;
Matches 15; Conservative 6; Mismatches 18; Indels 3; Gaps 1;

QY 3 VFSAEPSPPQTRSRMVFAHRLREVGDEFRSHLNSTDDADE 44
DB 83 VVSAADGPMPTQTRHILLS--RQGVPIVVFVFLNKADMVDD 121
RESULT 8
US-09-540-236-1942
; Sequence 1942, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARI
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 1942
; LENGTH: 373
; TYPE: PRT
; ORGANISM: M. catarrhalis
US-09-540-236-1942
Query Match 23.8%; Score 54.5; DB 4; Length 373;
Best Local Similarity 33.3%; Pred. No. 17;
Matches 14; Conservative 7; Mismatches 18; Indels 3; Gaps 1;
QY 3 VFSAEPSPPQTRSRMVFAHRLREVGDEFRSHLNSTDDADE 44
DB 113 VVSATDGPMPQTRHILLS--RQGVPIVVFVFLNKADMVDD 151
RESULT 9
US-09-140-466-4
; Sequence 4, Application US/09140466
; Patent No. 6268160
; GENERAL INFORMATION:
; APPLICANT: CLOUGH, BARBARA
; APPLICANT: PREISER, PETER
; APPLICANT: WILSON, ROBERT
; TITLE OF INVENTION: AN EF-TU PROTEIN ENCODED ON THE PLASTID DNA OF THE
; TITLE OF INVENTION: MALARIA PARASITE AND PROTEIN SYNTHESIS INHIBITORS
; FILE REFERENCE: N68837B GCW PJC DP
; CURRENT APPLICATION NUMBER: US/09/140,466
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: US 60/056,246
; EARLIER FILING DATE: 1997-08-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 4
; LENGTH: 408
; TYPE: PRT
; ORGANISM: Cryptomonas phi
US-09-140-466-4
Query Match 23.8%; Score 54.5; DB 3; Length 408;
Best Local Similarity 35.7%; Pred. No. 19;
Matches 15; Conservative 6; Mismatches 18; Indels 3; Gaps 1;
QY 3 VFSAEPSPPQTRSRMVFAHRLREVGDEFRSHLNSTDDADE 44
DB 104 VCSAAGPMPQTRHILLA---RQGVPIVVFVFLNKADMVDD 142
RESULT 10
US-08-599-480-4
; Sequence 4, Application US/08599480
; Patent No. 5753459
; GENERAL INFORMATION:
; APPLICANT: Bianco, David R.
; APPLICANT: Miller, James N.
; APPLICANT: Lovett, Michael A.

APPLICANT: Champion, Cheryl I.
APPLICANT: Tempst, Paul J.
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF A
TITLE OF INVENTION: T. Pallidum RARE OUTER MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/599,480
FILING DATE: 23-JAN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Learn, June M.
REGISTRATION NUMBER: 31,238
REFERENCE/DOCKET NUMBER: 07419/018001 (CIP of 016001)
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 678-5070
TELEFAX: (619) 678-5099
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 242 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-599-480-4

Query Match 23.6%; Score 54; DB 1; Length 242;
Best Local Similarity 38.2%; Pred. No. 12;
Matches 13; Conservative 4; Mismatches 17; Indels 0; Gaps 0;

QY 8 PSFPPTRRSMVFARHLREVGDPSRHLNSTDD 41
DB 172 PTHIQTSRYLGSAGHL3FVGFRTSPSERVDD 205

RESULT 11

US-09-252-991A-32486
Sequence 32486, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 32486
LENGTH: 267
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32486

Query Match 23.6%; Score 54; DB 4; Length 267;
Best Local Similarity 46.7%; Pred. No. 14;
Matches 14; Conservative 2; Mismatches 12; Indels 2; Gaps 1;

DB 7 ABPDPPQLQVARVFQR--RVPGQRCRCRH 34

RESULT 12

US-09-187-999-15
Sequence 15, Application US/09187999A
Patent No. 6482646
GENERAL INFORMATION:
APPLICANT: Gindullis, Frank
APPLICANT: Meier, Iris
TITLE OF INVENTION: Plant Proteins that Interact with Nuclear Matrix
TITLE OF INVENTION: Proteins and Function as Transcriptional Activators
FILE REFERENCE: CL-1321
CURRENT APPLICATION NUMBER: US/09/187,999A
CURRENT FILING DATE: 1998-11-06
NUMBER OF SEQ ID NOS: 39
SOFTWARE: Microsoft Office 97
SEQ ID NO 15
LENGTH: 339
TYPE: PRT
ORGANISM: Lycopersicon esculentum
US-09-187-999-15

Query Match 23.6%; Score 54; DB 4; Length 339;
Best Local Similarity 31.8%; Pred. No. 18;
Matches 14; Conservative 8; Mismatches 16; Indels 6; Gaps 2;

QY 7 EPSPPPTRR-----SMVFARHLREVGDPE-FSRHLNSTDDADE 44
DB 21 EPAPITSSGSLSPSPYLVNKHRRGGGEAFANRKLGLGEAEQ 64

RESULT 13

US-09-248-796A-25524
Sequence 25524, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
CURRENT FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 25524
LENGTH: 474
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-25524

Query Match 23.6%; Score 54; DB 4; Length 474;
Best Local Similarity 45.5%; Pred. No. 27;
Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 7 EPSPPPTRRSMVFARHLREV 28

DB 112 EPKPFPTKDDIVDKRLATTG 133

RESULT 14

US-09-140-466-5
Sequence 5, Application US/09140466
Patent No. 6288160
GENERAL INFORMATION:
APPLICANT: CLOUGH, BARBARA
APPLICANT: PREISER, PETER
APPLICANT: WILSON, ROBERT
TITLE OF INVENTION: AN FF-TM PROTEIN ENCODED ON THE PLASTID DNA OF THE


```
; TITLE OF INVENTION: EFFECTIVE AS ANTI-MALARIAL COMPOUNDS
; FILE REFERENCE: N68837B GCM BJC DP
; CURRENT APPLICATION NUMBER: US/09/140,466
; CURRENT FILING DATE: 1998-08-26
; EARLIER APPLICATION NUMBER: US 60/056,246
; EARLIER FILING DATE: 1997-08-28
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Cyanophora paradoxa
US-09-140-466-5

Query Match      23.4%; Score 53.5; DB 3; Length 409;
Best Local Similarity 35.7%; Pred No. 26;
Matches 15; Conservative 5; Mismatches 19; Indels 3; Gaps 1;

QY 3 VFSAPSPFPQTRSMVFARHLREVGDFRSRHLNSTDDADE 44
   ||||| : : : : :
DB 105 VVSAADGMPQTRHILIA--KQGVDPNMVFLNKEDQIDD 143

RESULT 15
US-09-252-991A-32205
; Sequence 32205, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32205
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32205

Query Match      23.4%; Score 53.5; DB 4; Length 433;
Best Local Similarity 33.3%; Pred No. 28;
Matches 19; Conservative 4; Mismatches 17; Indels 17; Gaps 2;

QY 1 AGVFSAPES-----PPQTRSMVFARHLREVGDFR-----SRHLNSTD 40
   ||||| : : : : :
DB 59 AGETSGEPQESGADRVFPQGRGAAVHARHLGLVDRDQGRGQPRRLPRAIHRRAGD 115

Search completed: November 10, 2004, 13:44:02
Job time : 15.1986 secs
```

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 16:36:12 ; Search time 47.2677 Seconds
(without alignments)
328.807 Million cell updates/sec

Title: US-10-092-750-29
Perfect score: 229
Sequence: 1 AGVFAEPPFPQTRSMVF.....REVGEFRSLNSTDDADE 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	229	100.0	44	14	US-10-092-750-29
2	155	67.7	353	15	US-10-309-290-118
3	155	67.7	429	16	US-10-471-450-8
4	63	27.5	415	16	US-10-437-963-121066
5	61	26.6	77	17	US-10-425-115-237269
6	61	26.6	414	17	US-10-425-115-240881
7	61	26.6	430	16	US-10-767-701-40545
8	61	26.6	436	16	US-10-437-963-178354
9	60.5	26.4	397	15	US-10-282-122A-52051
10	59	25.8	266	15	US-10-424-599-145647
11	58.5	25.5	282	15	US-10-282-122A-57410
12	57.5	25.1	209	15	US-10-282-122A-47669
13	57.5	25.1	235	15	US-10-282-122A-52391

14	57.5	25.1	380	15	US-10-282-122A-50564	Sequence 50564, A
15	57.5	25.1	396	15	US-10-282-122A-47841	Sequence 47841, A
16	57.5	25.1	396	15	US-10-282-122A-49112	Sequence 49112, A
17	57.5	25.1	396	15	US-10-282-122A-50484	Sequence 50484, A
18	57.5	25.1	396	15	US-10-282-122A-63962	Sequence 63962, A
19	57.5	25.1	397	15	US-10-282-122A-52408	Sequence 52408, A
20	57	24.9	122	14	US-10-067-974-14	Sequence 14, Appl
21	57	24.9	122	16	US-10-771-695-19	Sequence 19, Appl
22	57	24.9	250	9	US-09-738-628-5662	Sequence 5662, Ap
23	57	24.9	250	10	US-09-746-660A-36	Sequence 36, Appl
24	57	24.9	250	14	US-10-067-974-10	Sequence 10, Appl
25	57	24.9	250	16	US-10-771-695-16	Sequence 16, Appl
26	57	24.9	403	16	US-10-437-963-200906	Sequence 200906, A
27	56.5	24.7	396	15	US-10-282-122A-50797	Sequence 50797, A
28	56.5	24.7	396	15	US-10-282-122A-50956	Sequence 50956, A
29	56.5	24.7	396	15	US-10-282-122A-62675	Sequence 62675, A
30	56.5	24.7	396	15	US-10-282-122A-64421	Sequence 64421, A
31	56	24.5	194	16	US-10-767-701-57892	Sequence 57892, A
32	56	24.5	441	15	US-10-424-599-161614	Sequence 161614, A
33	56	24.5	441	15	US-10-424-599-161615	Sequence 161615, A
34	56	24.5	497	17	US-10-425-115-232275	Sequence 232275, A
35	56	24.5	518	15	US-10-425-114-70141	Sequence 70141, A
36	55.5	24.2	394	15	US-10-282-122A-65391	Sequence 65391, A
37	55.5	24.2	394	15	US-10-282-122A-65414	Sequence 65414, A
38	55.5	24.2	394	15	US-10-282-122A-65628	Sequence 65628, A
39	55.5	24.2	394	15	US-10-282-122A-65640	Sequence 65640, A
40	55.5	24.2	395	14	US-10-369-493-18496	Sequence 18496, A
41	55.5	24.2	395	15	US-10-282-122A-46036	Sequence 46036, A
42	55.5	24.2	396	15	US-10-282-122A-61924	Sequence 61924, A
43	55.5	24.2	397	9	US-09-815-242-11995	Sequence 11995, A
44	55.5	24.2	397	9	US-09-815-242-12002	Sequence 12002, A
45	55.5	24.2	397	15	US-10-282-122A-66623	Sequence 66623, A

ALIGNMENTS

RESULT 1
US-10-092-750-29
; Sequence 29, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-29

Query Match	100.0%	Score 229;	DB 14;	Length 44;
Best Local Similarity	100.0%	Pred. No. 1.6e-23;		
Matches 44;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AGVFAEPPFPQTRSMVFARHLREVGEFRSLNSTDDADE	44	
Db	1	AGVFAEPPFPQTRSMVFARHLREVGEFRSLNSTDDADE	44	

RESULT 2
US-10-309-290-118
; Sequence 118, Application US/10309290
; Publication No. US20040023241A1
; GENERAL INFORMATION:

```

; APPLICANT: Alsobrook II, John P.
; APPLICANT: Anderson, David W.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Chellakuru, Rajeev A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Gerlach, Valerie L.
; APPLICANT: Gorman, Linda
; APPLICANT: Gould-Rothberg, Bonnie E.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Jeffers, Michael E.
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Miller, Charles E.
; APPLICANT: Murphey, Ryan
; APPLICANT: Paturajan, Veera
; APPLICANT: Peyman, John A.
; APPLICANT: Rastelli, Luca
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Shency, Suresh G.
; APPLICANT: Smithson, Glenda
; APPLICANT: Starling, Gary
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Zhong, Haihong
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-502A
; CURRENT APPLICATION NUMBER: US/10/309,290
; CURRENT FILING DATE: 2002-12-02
; PRIOR APPLICATION NUMBER: 60/336,600
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/338,285
; PRIOR FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/341,346
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/341,477
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 60/341,540
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 60/342,592
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: 60/344,297
; PRIOR FILING DATE: 2001-12-27
; PRIOR APPLICATION NUMBER: 60/344,903
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 60/373,288
; PRIOR FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: 60/380,981
; PRIOR FILING DATE: 2002-05-15
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 274
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 118
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-290-118

Query Match 67.7%; Score 155; DB 15; Length 353;
Best Local Similarity 100.0%; Pred. No. 1.9e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 TRSMVFARHLREVGDFPRSHLNSTDDAD 43
Db 236 TRSMVFARHLREVGDFPRSHLNSTDDAD 265

RESULT 3
US-10-471-450-8
; Sequence 8, Application US/10471450
; Publication No. US20040152877A1

; APPLICANT: Alsobrook II, John P.
; APPLICANT: Anderson, David W.
; APPLICANT: Boldog, Ferenc L.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Chellakuru, Rajeev A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Gerlach, Valerie L.
; APPLICANT: Gorman, Linda
; APPLICANT: Gould-Rothberg, Bonnie E.
; APPLICANT: Guo, Xiaojia
; APPLICANT: Jeffers, Michael E.
; APPLICANT: Ji, Weizhen
; APPLICANT: Li, Li
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Miller, Charles E.
; APPLICANT: Murphey, Ryan
; APPLICANT: Paturajan, Veera
; APPLICANT: Peyman, John A.
; APPLICANT: Rastelli, Luca
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Shency, Suresh G.
; APPLICANT: Smithson, Glenda
; APPLICANT: Starling, Gary
; APPLICANT: Taupier, Raymond J.
; APPLICANT: Voss, Edward Z.
; APPLICANT: Zhong, Haihong
; APPLICANT: Zhong, Mei
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-502A
; CURRENT APPLICATION NUMBER: US/10/309,290
; CURRENT FILING DATE: 2002-12-02
; PRIOR APPLICATION NUMBER: 60/336,600
; PRIOR FILING DATE: 2001-12-05
; PRIOR APPLICATION NUMBER: 60/338,285
; PRIOR FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: 60/341,346
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/341,477
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 60/341,540
; PRIOR FILING DATE: 2001-12-17
; PRIOR APPLICATION NUMBER: 60/342,592
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: 60/344,297
; PRIOR FILING DATE: 2001-12-27
; PRIOR APPLICATION NUMBER: 60/344,903
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 60/373,288
; PRIOR FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: 60/380,981
; PRIOR FILING DATE: 2002-05-15
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 274
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 118
; LENGTH: 353
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-309-290-118

Query Match 67.7%; Score 155; DB 15; Length 353;
Best Local Similarity 100.0%; Pred. No. 1.9e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 TRSMVFARHLREVGDFPRSHLNSTDDAD 43
Db 236 TRSMVFARHLREVGDFPRSHLNSTDDAD 265

RESULT 4
US-10-437-963-121066
; Sequence 121066, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963

```

```

; GENERAL INFORMATION:
; APPLICANT: JACKSON, Jennifer L.; BAUGHN, Mariah R.;
; APPLICANT: SWARNAKAR, Anita; YUE, Henry;
; APPLICANT: ELLIOTT, Vicki S.; BURFORD, Neil;
; APPLICANT: DING, Li; TANG, Y. Tom;
; APPLICANT: LEE, Soo Yeun; AZIMZAI, Valda;
; APPLICANT: CHAWLA, Narinder K.; GIETZEN, Kimberly J.;
; APPLICANT: GRIFFIN, Jennifer A.; LAL, Preeti G.;
; APPLICANT: YANG, Junming; BOROWSKY, Mark L.;
; APPLICANT: RICHARDSON, Thomas W.; YUE, Huibin;
; APPLICANT: BECHA, Shanya; FORSYTHE, Ian J.;
; APPLICANT: JONES, Karen Anne; WARREN, Bridget;
; APPLICANT: THANGAVELU, Kavitha; HONCHHELL, Cynthia;
; APPLICANT: JOLLEY, Helen E.; HAFALIA, April J.A.; Huijun Z. Ring
; TITLE OF INVENTION: NUCLEIC ACID ASSOCIATED PROTEINS
; FILE REFERENCE: PF-0917 USN
; CURRENT APPLICATION NUMBER: US/10/471,450
; CURRENT FILING DATE: 2003-09-10
; PRIOR APPLICATION NUMBER: PCT/US02/07869
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: US 60/276,857
; PRIOR FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/285,489
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/285,556
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/288,700
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/288,646
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/290,510
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: US 60/290,369
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/332,426
; PRIOR FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 7493913CD1
US-10-471-450-8

Query Match 67.7%; Score 155; DB 16; Length 429;
Best Local Similarity 100.0%; Pred. No. 2.3e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 14 TRSMVFARHLREVGDFPRSHLNSTDDAD 43
Db 236 TRSMVFARHLREVGDFPRSHLNSTDDAD 265

RESULT 4
US-10-437-963-121066
; Sequence 121066, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963

```

```
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 121066
; LENGTH: 415
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_24127C.1.pep
US-10-437-963-121066
```

```
Query Match 27.5%; Score 63; DB 16; Length 415;
Best Local Similarity 46.9%; Pred. No. 6.5;
Matches 15; Conservative 2; Mismatches 11; Indels 4; Gaps 1;

QY 3 VFSPFPFP-----QTRRSVMVFARHLREVGDE 30
Db 30 VLFVSPFPAYISGYKNRFQNFIFKHLREMGDE 61
```

```
RESULT 5
US-10-425-115-237269
; Sequence 237269, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 237269
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_147975C.1.pep
US-10-425-115-237269
```

```
Query Match 26.6%; Score 61; DB 17; Length 77;
Best Local Similarity 50.0%; Pred. No. 1.7;
Matches 14; Conservative 2; Mismatches 8; Indels 4; Gaps 1;

QY 7 EPSFPFP-----QTRRSVMVFARHLREVGDE 30
Db 7 EPSFPAYISGYKNRFQNFIFKHLREMGDE 34
```

```
RESULT 6
US-10-425-115-240881
; Sequence 240881, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 240881
; LENGTH: 414
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_151262C.1.pep
US-10-425-115-240881
```

```
Query Match 26.6%; Score 61; DB 17; Length 414;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 14; Conservative 2; Mismatches 8; Indels 4; Gaps 1;

QY 7 EPSFPFP-----QTRRSVMVFARHLREVGDE 30
Db 33 EPSFPAYISGYKNRFQNFIFKHLREMGDE 60
```

```
RESULT 7
US-10-767-701-40545
; Sequence 40545, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 40545
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C82521_1.pep
US-10-767-701-40545
```

```
Query Match 26.6%; Score 61; DB 16; Length 430;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 14; Conservative 2; Mismatches 8; Indels 4; Gaps 1;

QY 7 EPSFPFP-----QTRRSVMVFARHLREVGDE 30
Db 27 EPSFPAYISGYKNRFQNFIFKHLREMGDE 54
```

```
RESULT 8
US-10-437-963-178354
; Sequence 178354, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 178354
; LENGTH: 436
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_75918C.1.pep
US-10-437-963-178354
```

```
Query Match 26.6%; Score 61; DB 16; Length 436;
Best Local Similarity 50.0%; Pred. No. 13;
Matches 14; Conservative 2; Mismatches 8; Indels 4; Gaps 1;

QY 7 EPSFPFP-----QTRRSVMVFARHLREVGDE 30
```

Db 33 EBPSPAYISGVKNRQNFQIKHLRENGDE 60
||||| :|||:||||

RESULT 9

US-10-282-122A-52051
; Sequence 52051, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 52051
; LENGTH: 397
; TYPE: PRT
; ORGANISM: Clostridium acetobutylicum

US-10-282-122A-52051
Query Match 26.4%; Score 60.5; DB 15; Length 397;
Best Local Similarity 38.1%; Pred. No. 13;
Matches 16; Conservative 4; Mismatches 19; Indels 3; Gaps 1;

QY 3 VFSARPSFPQTRSRMVFAHRLREYGVDFRSRHLNSTDDADE 44
||| :|||:||||

Db 105 VVSAADGMPQTRHEILLA---SRVGEYIVVFLNKADQVDD 143
||| :|||:||||

RESULT 10

US-10-424-599-145647
; Sequence 145647, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with

US-10-424-599-145647
Query Match 25.8%; Score 59; DB 15; Length 266;
Best Local Similarity 35.6%; Pred. No. 13;
Matches 16; Conservative 5; Mismatches 14; Indels 10; Gaps 2;

QY 10 PFPQ-----TRSRMVFAHRLREYGVDFRSRHLNSTDDADE 44
||| :|||:||||

Db 31 PFPNSNASSPFTVRASVFAQWPHLYRFSKDKDDHSDDDGDE 75
||| :|||:||||

RESULT 11

US-10-282-122A-57410
; Sequence 57410, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57410
; LENGTH: 282
; TYPE: PRT

```
; ORGANISM: Enterococcus faecium
US-10-282-122A-57410

Query Match      25.1%; Score 58.5; DB 15; Length 282;
Best Local Similarity 35.7%; Pred. No. 17;
Matches 15; Conservative 7; Mismatches 17; Indels 3; Gaps 1;

QY 3 VFSAPSPFPQTRSRMVFAHRLREVGDEFRRSHLNSTDDADE 44
Db 17 VVSATDGMFPQTRHILLIA---RQGVKYLIVFLNKVDLVD 55

RESULT 12
US-10-282-122A-47669
; Sequence 47669, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 47669
; LENGTH: 209
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47669

Query Match      25.1%; Score 57.5; DB 15; Length 209;
Best Local Similarity 38.1%; Pred. No. 16;
Matches 16; Conservative 5; Mismatches 18; Indels 3; Gaps 1;

QY 3 VFSAPSPFPQTRSRMVFAHRLREVGDEFRRSHLNSTDDADE 44
Db 105 VCSAADGMPQTRHILLIA---RQGVPIIVFLNKCDMVD 143

RESULT 13
US-10-282-122A-52391
```

```
; Sequence 52391, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 52391
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Clostridium botulinum
US-10-282-122A-52391

Query Match      25.1%; Score 57.5; DB 15; Length 235;
Best Local Similarity 35.7%; Pred. No. 18;
Matches 15; Conservative 5; Mismatches 19; Indels 3; Gaps 1;

QY 3 VFSAPSPFPQTRSRMVFAHRLREVGDEFRRSHLNSTDDADE 44
Db 105 VVSADGMPQTRHILLIA---SRGVQYIVVFLNKADQVDD 143

RESULT 14
US-10-282-122A-50564
; Sequence 50564, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
```

```

/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITRA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 50564
/ LENGTH: 380
/ TYPE: PRT
/ ORGANISM: Burkholderia mallei
/ US-10-282-122A-50564

```

```
Query Match      25.1%; Score 57.5; DB 15; Length 380;
Best Local Similarity 38.1%; Pred. No. 32;
Matches 16; Conservative 5; Mismatches 18; Indels 3; Gaps 1;
```

```

RESULT 15
US-10-282-122A-47841
; Sequence 47841, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Heselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347

```

```

; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent version 3.1
; SEQ ID NO 47841
; LENGTH: 396
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47841

Query Match                25.1%; Score 57.5; DB 15; Length 396;
Best Local Similarity      38.1%; Pred. No. 34;
Matches 16; Conservative   5; Mismatches 18; Indels    3; Gaps 1

QY       3 VFSAPSPFPQTGRSMVFARHLREVGDFRSRLHNSLTDDADE 44
          |||::|||::|||::|||::|||::|||::|||::|||::
Db       105 VCSAADGPMQTEHHLLA---RQGVPIIVFLNKCWMDD 143

Search completed: November 11, 2004, 01:28:11
Job time : 48.3177 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:29:32 ; Search time 9.80311 Seconds
(without alignments)
431.857 Million cell updates/sec

Title: US-10-092-750-29
Perfect score: 229
Sequence: 1 AGVFAEPSPPQFRSMVF.....REVGFPSRHLNSTDDAD 44

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	155	67.7	383	2 T14795	hypothetical prote
2	60.5	26.4	396	2 D60663	translation elonga
3	60.5	26.4	397	2 H97285	elongation factor
4	59.5	26.0	301	2 T27351	hypothetical prote
5	59	25.8	470	2 T45962	hypothetical prote
6	58.5	25.5	394	1 ERYMTS	translation elonga
7	58.5	25.5	396	2 S55281	translation elonga
8	58.5	25.5	396	2 A82532	translation elonga
9	57.5	25.1	396	2 G87143	translation elonga
10	57.5	25.1	397	2 S23909	translation elonga
11	57.5	25.1	406	1 S00229	translation elonga
12	57.5	25.1	406	1 S17146	translation elonga
13	57.5	25.1	406	1 S29293	translation elonga
14	57.5	25.1	776	2 A87330	sensory box histid
15	56.5	24.7	391	2 A82818	elongation factor
16	56.5	24.7	391	2 C97596	elongation factor
17	56.5	24.7	391	2 AB2816	elongation factor
18	56.5	24.7	391	2 D97594	elongation factor
19	56.5	24.7	391	2 AB3346	protein translatio
20	56.5	24.7	396	2 A44795	translation elonga
21	56.5	24.7	396	2 T44381	translation elonga
22	56.5	24.7	397	2 I40216	translation elonga
23	56.5	24.7	406	2 AH3344	protein translatio
24	55.5	24.2	394	2 T10168	translation elonga
25	55.5	24.2	394	2 A81235	translation elonga
26	55.5	24.2	394	2 A81234	translation elonga
27	55.5	24.2	395	2 E66957	elongation factor
28	55.5	24.2	397	2 F83111	elongation factor
29	55.5	24.2	509	2 A53741	transcription fact

30 55 24.0 250 2 B40626 probable 2,3-dihyd
31 55 24.0 827 2 A69793 conserved hypothet
32 54.5 23.8 399 2 A81392 translation elonga
33 54.5 23.8 405 2 E75533 translation elonga
34 54.5 23.8 405 2 C60663 translation elonga
35 54.5 23.8 409 2 S73208 translation elonga
36 54.5 23.8 452 2 S68466 translation elonga
37 54.5 23.8 452 2 S62768 translation elonga
38 54.5 23.8 452 2 S62767 translation elonga
39 54.5 23.8 849 2 T49342 fzz related protei
40 54 23.6 242 2 H71297 probable outer mem
41 54 23.6 844 2 T37690 hypothetrical prote
42 53.5 23.4 394 2 C71672 translation elonga
43 53.5 23.4 395 2 AD1406 translation elonga
44 53.5 23.4 395 2 AD1782 translation elonga
45 53.5 23.4 396 2 A87403 translation elonga

ALIGNMENTS

RESULT 1

T14795
hypothetical protein DKFZp434E171.1 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C/Accession: T14795
R/Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, August 1999
A/Reference number: Z18183
A/Accession: T14795
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-383 <POU>
A/Cross-references: EMBL:AL110285
A/Experimental source: adult testis; clone DKFZp434E171
C/Genetics:
A/Note: DKFZp434E171.1

Query Match 67.7%; Score 155; DB 2; Length 383;
Best Local Similarity 100.0%; Pred. No. 1.2e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 TRRSVMFARHLREVGVDFPSRHLNSTDDAD 43
DB 239 TRRSVMFARHLREVGVDFPSRHLNSTDDAD 268

RESULT 2

D60663
translation elongation factor EF-Tu - Pseudomonas cepacia
C/Species: Pseudomonas cepacia
C/Date: 28-Apr-1993 #sequence_revision 28-Apr-1993 #text_change 09-Jul-2004
C/Accession: D60663
R/Ludwig, W.; Weizenegger, M.; Betzl, D.; Leidel, E.; Lenz, T.; Ludvigsen, A.; Moellenhof
Arch. Microbiol. 153, 241-247, 1990
A/Title: Complete nucleotide sequences of seven eubacterial genes coding for the elongat
A/Reference number: A60663; MUID:90240875; PMID:2110445
A/Accession: D60663
A/Status: not compared with conceptual translation
A/Molecule type: DNA
A/Residues: 1-396 <LUD>
A/Cross-references: UNIPROT:P33167
C/Superfamily: translation elongation factor Tu; translation elongation factor Tu homolog
C/Keywords: GTP binding; nucleotide binding; p-loop; protein biosynthesis
F/13-139/Domain: translation elongation factor Tu homology <ETU>
F/19-26/Region: nucleotide-binding motif A (P-loop)
F/136-139/Region: GTP-binding NKXD motif
F/174-176/Region: GTP-binding SAK/L motif
F/25,26,62,136,137,139,174/Binding site: Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser) #stat

Query Match 26.4%; Score 60.5; DB 2; Length 396;
Best Local Similarity 38.1%; Pred. No. 3.6;

Matches 16; Conservative 5; Mismatches 18; Indels 3; Gaps 1;
 QY 3 VFSAPSPFPQTRSRMVFAHRLREVGDFRSRHLNSTDDADE 44
 Db 105 VCSAADGPMFPQTRHEHLLA---RQGVGPYIIIVFLNKCDSDVD 143

RESULT 3
 H97285
 elongation factor Tu (Ef-Tu) [imported] - Clostridium acetobutylicum
 C:Species: Clostridium acetobutylicum
 C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Jul-2004
 C:Accession: H97285
 R:Rolling, J.; Berton, G.; Onelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: H97285
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-397 <KUR>
 A:Cross-references: UNIPROT:Q97EH5; GB:AE001437; PIDN:AAK81075.1; PID:gi5026203; GSPDB:G0000023
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC3136
 C:Superfamily: translation elongation factor Tu; translation elongation factor Tu homolog

Query Match 26.4%; Score 60.5; DB 2; Length 397;
 Best Local Similarity 38.1%; Pred. No. 3.6;
 Matches 16; Conservative 4; Mismatches 19; Indels 3; Gaps 1;
 QY 3 VFSAPSPFPQTRSRMVFAHRLREVGDFRSRHLNSTDDADE 44
 Db 105 VVSAADGPMFPQTRHEHLLA---SRVGVEYIVVFLNKADQDVD 143

RESULT 4
 T27351
 hypothetical protein Y70C5C.1 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T27351
 R:Steward, C.
 submitted to the EMBL Data Library, September 1997
 A:Reference number: Z20351
 A:Accession: T27351
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-301 <WIL>
 A:Cross-references: UNIPROT:O62499; EMBL:Z99283; PIDN:CAB16537.1; GSPDB:GN000023; CESP:Y70C5C
 A:Experimental source: clone Y70C5C
 C:Genetics:
 A:Gene: CESP:Y70C5C.1
 A:Map position: 5
 A:Introns: 113/1; 225/2

Query Match 26.0%; Score 59.5; DB 2; Length 301;
 Best Local Similarity 39.5%; Pred. No. 3.6;
 Matches 15; Conservative 5; Mismatches 17; Indels 1; Gaps 1;
 QY 5 SAEPSPFPQTRSRMVFAHRLREVGDFRSRHLNSTDD 41
 Db 33 SAAPSSRLFRNEHILRLRLQNNQNGDEYIYRHLQTHD 70

RESULT 5
 T45962
 hypothetical protein F7J8.200 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C>Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004
 C:Accession: T45962
 R:Bevan, M.; Zimmermann, W.; Gruenisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Lem

submitted to the Protein Sequence Database, January 2000
 A:Reference number: Z23018
 A:Accession: T45962
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-470 <BEV>
 A:Cross-references: UNIPROT:Q9LFB4; EMBL:AL137189
 A:Experimental source: cultivar Columbia; BAC clone F7J8
 C:Genetics:
 A:Map position: 5
 A:Introns: 116/2; 138/3; 162/2; 201/3; 226/2; 250/1; 275/1; 288/2; 307/2; 379/3
 A:Note: F7J8.200

Query Match 25.8%; Score 59; DB 2; Length 470;
 Best Local Similarity 50.0%; Pred. No. 6.8;
 Matches 14; Conservative 2; Mismatches 8; Indels 4; Gaps 1;
 QY 7 EPSPPFPQT---RRSMVFARHLREVGDE 30
 Db 111 EPSPPFAYVSGYKXRFQNFYRLREMGDE 138

RESULT 6
 EFYVTS
 translation elongation factor EF-Tu - Mycoplasma gallisepticum
 C:Species: Mycoplasma gallisepticum
 C>Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 09-Jul-2004
 C:Accession: S14910; S06704
 R:Inamine, J.M.; Loechel, S.; Hu, P.C.
 Nucleic Acids Res. 17, 10126, 1989
 A:Title: Nucleotide sequence of the tuf gene from Mycoplasma gallisepticum.
 A:Reference number: S14910; MUID:90098800; PMID:2602129
 A:Accession: S14910
 A:Molecule type: DNA
 A:Residues: 1-394 <INA>
 A:Cross-references: UNIPROT:P18906; EMBL:X16462; NID:g44292; PIDN:CAA34482.1; PID:g44293
 C:Genetics:
 A:Gene: tuf
 A:Genetic code: SGC3
 C:Superfamily: translation elongation factor Tu; translation elongation factor Tu homolog
 C:Keywords: GTP binding; nucleotide binding; P-loop; protein biosynthesis
 F:13-139/Domain: translation elongation factor Tu homology <ETU>
 F:19-26/Region: nucleotide-binding motif A (P-loop)
 F:136-139/Region: GTP-binding NKXD motif
 F:174-176/Region: GTP-binding SAK/L motif
 F:25,26,62,136,137,139,174/Binding site: Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser) #stat

Query Match 25.5%; Score 58.5; DB 1; Length 394;
 Best Local Similarity 40.5%; Pred. No. 6.5;
 Matches 17; Conservative 4; Mismatches 18; Indels 3; Gaps 1;
 QY 3 VFSAPSPFPQTRSRMVFAHRLREVGDFRSRHLNSTDDADE 44
 Db 105 VVSAADGPMFPQTRHEHLLA---RQGVGPYIIIVFLNKCDSDVD 143

RESULT 7
 S55281
 translation elongation factor EF-Tu - Stigmatella aurantiaca
 C:Species: Stigmatella aurantiaca
 C>Date: 23-Aug-1995 #sequence_revision 19-Oct-1995 #text_change 09-Jul-2004
 C:Accession: S55281; S52237
 R:Bremaud, L.; Fremaux, C.; Laalami, S.; Cenatiempo, Y.
 Nucleic Acids Res. 23, 1737-1743, 1995
 A:Title: Genetic and molecular analysis of the tRNA-tufB operon of the myxobacterium Stigmatella aurantiaca
 A:Reference number: S55281; MUID:95303606; PMID:7784178
 A:Accession: S55281
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-396 <BRE>
 A:Cross-references: UNIPROT:P42479; EMBL:X82820; NID:g509255; PIDN:CAA58029.1; PID:g50925
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1994
 C:Superfamily: translation elongation factor Tu; translation elongation factor Tu homolog

C;Keywords: GTP binding; nucleotide binding; P-loop
F;13-139/Domain: translation elongation factor Tu homology <ETU>
F;19-26/Region: nucleotide-binding motif A (P-loop)
F;136-139/Region: GTP-binding NKXD motif
F;174-176/Region: GTP-binding SAK/L motif
F;25,26,62,136,137,139,174/Binding site: Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser) #sta

Query Match 25.5%; Score 58.5; DB 2; Length 396;
Best Local Similarity 38.1%; Pred. No. 6.5; Mismatches 5; Indels 3; Gaps 1;
Matches 16; Conservative

QY 3 VFSAPSPFPQTRSRMVFAHRLREVGDFFSRHLSNSTDADA 44
Db 105 VCSAADGMPQTRHILLA---RQGVFPYIVVFLNKADMDVD 143

RESULT 9
A82532
translation elongation factor EF-Tu XF2640 XP2628 [imported] - Xylella fastidiosa (strain
C;Species: Xylella fastidiosa
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C;Accession: A82532; E82534
R;anonymus, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A;Reference number: A82515; MUID:20365717; PMID:10910347
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: A82532
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-396 <SIM>
A;Cross-references: UNIPROT:Q9P9Q9; GB:AE004071; GB:AE003849; NID:99107866; PIDN:AAF8543
A;Experimental source: strain 945C
A;Expression: E82534
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-396 <SI2>
A;Cross-references: GB:AB004069; GB:AB003849; NID:99107849; PIDN:AAF85425.1; GSPDB:GN001
A;Experimental source: strain 945C
R;Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, J.B.; Kitajima, J.P.; Krieger, J.B.; Kuramae, E.E.; Laigh
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A;Authors: Martins, E.M.F.; Matekuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
P.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A;Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

Query Match 25.5%; Score 58.5; DB 2; Length 396;
Best Local Similarity 38.1%; Pred. No. 6.5; Mismatches 5; Indels 3; Gaps 1;
Matches 16; Conservative

QY 3 VFSAPSPFPQTRSRMVFAHRLREVGDFFSRHLSNSTDADA 44
Db 105 VCSAADGMPQTRHILLA---RQGVFPYIVVFLNKADMDVD 143

RESULT 9
G87143
elongation factor EF-Tu [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae

C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C;Accession: G87143
R;Cole, S.T.; Eglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor
R.; Davies, R.M.; Devlin, K.; Duchoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: G87143
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-396 <STO>
A;Cross-references: UNIPROT:P30768; GB:AL450380; NID:913093557; PIDN:CAC30831.1; GSPDB:G1
C;Genetics:
A;Gene: tuf
C;Superfamily: translation elongation factor Tu; translation elongation factor Tu homolo

Query Match 25.1%; Score 57.5; DB 2; Length 396;
Best Local Similarity 35.7%; Pred. No. 8.9; Mismatches 7; Indels 3; Gaps 1;
Matches 15; Conservative

QY 3 VFSAPSPFPQTRSRMVFAHRLREVGDFFSRHLSNSTDADA 44
Db 107 VVAATDGPMPQTRHILLA---RQGVFPYIVVFLNKSDAVDD 145

RESULT 10
S23909
translation elongation factor EF-Tu.2 - Streptomyces ramocissimus
C;Species: Streptomyces ramocissimus
C;Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 09-Jul-2004
C;Accession: S23909
R;Wondt, L.P.; Rierveld, K.; Verdurmen, M.; van Haarlem, J.; van Wezei, G.P.; Vijgenboom,
submitted to the EMBL Data Library, June 1992
A;Description: Three tuf (-like) genes in Streptomyces ramocissimus.
A;Reference number: S23907
A;Accession: S23909
A;Molecule type: DNA
A;Residues: 1-397 <WON>
A;Cross-references: UNIPROT:P29543; EMBL:X67058; NID:947485; PIDN:CAA47443.1; PID:G58173;
C;Genetics:
A;Gene: tuf2
A;Start codon: GTG
C;Superfamily: translation elongation factor Tu; translation elongation factor Tu homolo
C;Keywords: GTP binding; nucleotide binding; P-loop; protein biosynthesis
F;13-141/Domain: translation elongation factor Tu homology <ETU>
F;19-26/Region: nucleotide-binding motif A (P-loop)
F;138-141/Region: GTP-binding NKXD motif
F;176-178/Region: GTP-binding SAK/L motif
F;25,26,64,138,139,141,176/Binding site: Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser) #stat

Query Match 25.1%; Score 57.5; DB 2; Length 397;
Best Local Similarity 35.7%; Pred. No. 8.9; Mismatches 7; Indels 3; Gaps 1;
Matches 15; Conservative

QY 3 VFSAPSPFPQTRSRMVFAHRLREVGDFFSRHLSNSTDADA 44
Db 107 VVAATDGPMPQTRHILLA---RQGVFPYIVVFLNKTDVDD 145

RESULT 11
S00229
translation elongation factor EF-Tu.A version 1 [validated] - Thermus aquaticus
N;Contains: GTPase (EC 3.6.1.-)
C;Species: Thermus aquaticus
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 09-Jul-2004
C;Accession: S00229; A27277
R;Kushiro, A.; Shimizu, M.; Tomita, K.I.
Eur. J. Biochem. 170, 93-98, 1987
A;Title: Molecular cloning and sequence determination of the tuf gene coding for the elor
A;Reference number: S00229; MUID:68082865; PMID:2826164
A;Accession: S00229

F:137-140/Region: GTP-binding NKXD motif	
F:175-177/Region: GTP-binding SAK/L motif	
F:95,26,63,137,138,140,175/Binding site: Mg-GTP (Lys, Thr, Asn, Lys, Asp, Ser) #sta	
Query Match	25.1%; Score 57.5; DB 1; Length 406;
Db	100 VCSAADGPMPTREHLLA---RQVGPAIVVFLNKKVDQVDD 138
Search completed:	November 10, 2004, 13:40:37
Job time :	9.80311 secs

Query Match 25.1%; Score 57.5; DB 1; Length 406;
Best Local Similarity 35.7%; Pred. No. 9.1;
Matches 15; Conservative 6; Mismatches 18; Indels 3; Gaps 1;

QY	3	VFSAEPSPFPQTRSMVFARHLREVGDEFRRSHLNSTDDADE	44
Dib	106	VVSAADGEMPQTRZHHLLA---RQVGVPYIVWFNMVCDMVDD	144

RESULT 14

A87350
sensory box histidine kinase/response regulator [imported] - *Caulobacter crescentus*

C/Species: *Caulobacter crescentus*
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C/Accession: A87330
R/Nierman, W.C.; Fealdilyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A/Title: Complete Genome Sequence of *Caulobacter crescentus*.
A/Reference number: A87249; MUID:21173698; PMID:11259547
A/Accession: A87330
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-776 <STOP>
A/Cross-references: UNIPROT:Q9AAE9; GB:AE005673; NID:g13421867; PIDN:AAK22637.1; GSPDB:G
C/Genetics:
A/Gene: CC0652

Query Match	25.1%	Score 57.5;	DB 2;	Length 776;
Best Local Similarity	44.7%	Pred. No. 19;		
Matches	17;	Conservative	3;	Mismatches 13;
				Indels 5;
				Gaps 2;

Qy 8 P S P F P Q --- T R R S M V F A R H L R E V G D E F R S R H - L N S T D 40
| : | | | | | | | | | | : | | | | |
D b 2 P T P F P O D I D A L R A S E A F F R L L T E O A G D V I S R H R L S G T D 39

RESULT 15

AC2818
elongation factor TU [imported] - Agrobacterium tumefaciens (strain C58, Duport)
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C/Accession: AC2818
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
erage, G.; Gillet, W.; Grant, C.; Guenthrner, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClellan,
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AC2818
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-391 <KUR>
A/Cross-references: UNIPROT:Q8UE16; GB:A5008688; PIDN:AAL42961.1; PID:gl7740420; GSPDB:G
A/Experimental source: strain C58 (Duport)
C/Genetics:
A/Gene: tuf
A/Map position: circular chromosome
C/Superfamily: translation elongation factor Tu; translation elongation factor Tu homolog

Query Match 24.7%; Score 56.5; DB 2; Length 391;
Best Local Similarity 38.1%; Pred. No. 12;
Matches 16: Conservative 4; Mismatches 19; Indels 3; Gaps 1;

QY 3 VFSAEPSFPQTRRSMVFARHILREVGVDEFRRSRHLNSTDDADE 44

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 53.7271 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-29

Perfect score: 229

Sequence: 1 AGVFAEPSPFPQTRBSWVF.....REVGDFRSHLSTDDADE 44

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	155	67.7	424	1	Q9Y295 homo sapien
2	155	67.7	429	2	Q7Z4N0
3	155	67.7	429	2	Q6EV56
4	155	67.7	429	2	AAL47681
5	155	67.7	429	2	AAL47681
6	136	59.4	141	2	Q6PJV1
7	136	59.4	141	2	AAL11044
8	135	59.0	429	1	Q8VH13
9	134	58.5	429	2	Q7YR55
10	134	58.5	429	2	Q7YR55
11	119	52.0	380	2	Q7T1N6
12	117	51.1	386	2	Q6EV53
13	79	34.5	201	2	Q99LA7
14	75	32.8	434	2	Q7VZ57
15	74	32.3	423	2	Q7K729
16	74	32.3	490	2	Q6W589
17	72	31.4	403	2	Q6EV52
18	67.5	29.5	254	2	Q6KMN9
19	66.5	29.0	254	2	Q6KMN6
20	66.5	29.0	254	2	Q6KMN3
21	66.5	29.0	254	2	Q6KMN4
22	66.5	29.0	278	2	Q6EZV5
23	66	28.8	424	1	Q6U2_CAEEL
24	66	27.9	424	2	Q6EV55
25	63.5	27.7	232	2	Q6UE25
26	63.5	27.7	232	2	AAR25433
27	63.5	27.7	233	2	Q6UE27
28	63.5	27.7	233	2	AAR25431
29	63.5	27.7	254	2	Q6KH86
30	63.5	27.7	254	2	Q6KHU6
31	63.5	27.7	254	2	Q6KHJ7

ALIGNMENTS

RESULT 1

OFU2 HUMAN STANDARD; PRT: 424 AA.
 AC Q9Y2G5; Q8WU6; Q9BQS4; Q9BQS5; Q9UFY3;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE GDP-fucose protein O-fucosyltransferase 2 precursor (BC 2.4.1.221)
 DE (Peptide O-fucosyltransferase) (O-FucT-2).
 GN Name=POPUT2; Synonyms=C21orf80, KIAA0958;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RC TISSUE=Brain;
 RX MEDLINE=99246063; PubMed=10231032;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hiroseawa M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIII. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";
 RL DNA Res. 6:63-70(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A; B AND C).
 RC TISSUE=Testis;
 RX Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RT "C21orf80, a new gene from the 21q22.3 region: using RNAi to get insight into the function.";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM B).
 RC TISSUE=Testis;
 RX Poustka A., Klein M., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20289799; PubMed=10830953;
 RA Rattori M., Fujiyama A., Taylor I.D., Watanabe H., Yada T., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y., Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D., Reichwald K., Rump J., Schillabel M., Schudy A., Zimmermann W., Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S., Shintani A., Sasaki T., Nagamine K., Mitsuyma S., Antonarakis S.E., Minoshima S., Shimizu N., Nordieck G., Hornischer K., Brandt P., Schaefer M., Schoen O., Desario A., Reichelt J., Kauer G., Bloeker H., Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E., Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., Yaspo M.-L.;
 RT "The DNA sequence of human chromosome 21.";
 RL Nature 405:311-319(2000).
 CC -!- FUNCTION: Catalyzes the reaction that attaches fucose through an O-glycosidic linkage to a conserved serine or threonine residue in

32 63.5 27.7 254 2 Q6KHJ8 lactobacill
 33 63.5 27.7 254 2 Q6KHJ8 lactobacill
 34 63.5 27.7 254 2 Q6KHJ8 lactobacill
 35 63.5 27.7 254 2 Q6KHJ8 lactobacill
 36 63.5 27.7 254 2 Q6KHJ8 lactobacill
 37 63.5 27.7 254 2 Q6KHJ8 lactobacill
 38 63.5 27.7 254 2 Q6KHJ8 lactobacill
 39 63.5 27.7 254 2 Q6KHJ8 lactobacill
 40 63.5 27.7 254 2 Q6KHJ8 lactobacill
 41 63.5 27.7 254 2 Q6KHJ8 lactobacill
 42 63.5 27.7 254 2 Q6KHJ8 lactobacill
 43 63.5 27.7 277 2 Q6KHJ8 lactobacill
 44 63.5 27.7 395 1 EFTU LACPL
 45 63 27.5 415 2 Q6KHJ8 lactobacill

CC EGF domains (By similarity).

CC -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-

CC beta-L-fucose to the serine hydroxy group of a protein acceptor.

CC -!- PATHWAY: Glycosylation.

CC -!- ALTERNATIVE PRODUCTS:

CC Event-Alternative splicing; Named isoforms=3;

CC Name=A;

CC IsoId=Q9Y2G5-1; Sequence=Displayed;

CC Name=B;

CC IsoId=Q9Y2G5-2; Sequence=VSP_003833, VSP_003834;

CC Name=C;

CC IsoId=Q9Y2G5-3; Sequence=VSP_003832;

CC -!- SIMILARITY: Belongs to the glycosyltransferase family 68.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC DR EMBL; AB023175; BAA76802.1; ALT_INIT.

CC DR EMBL; AJ302080; CAC24557.1; -.

CC DR EMBL; AJ302079; CAC24556.1; -.

CC DR EMBL; AY066015; AAL47681.1; -.

CC DR EMBL; AL110285; CAB53715.1; ALT_INIT.

CC DR EMBL; AL163301; CAB90496.1; ALT_INIT.

CC DR Genew; HGNC:14683; POPUT2.

CC KW Alternative splicing; Fucose metabolism; Glycoprotein;

CC Glycosyltransferase; Signal; Transferase.

FT SIGNAL 1 21

FT CHAIN 22 424

FT 2

FT CARBOHYD 189

FT CARBOHYD 209

FT CARBOHYD 259

FT VARSPLIC 380 424

FT CLLV -> ETPGEGCPHQIQKVGCGGLDLRDVYCAG

FT (in isoform C).

FT /FTId=VSP_003832.

FT RC> SS (in isoform B).

FT /FTId=VSP_003833.

FT Missing (in isoform B).

FT /FTId=VSP_003834.

FT SQ SEQUENCE 424 AA; 48895 MW; 60BA687F0DFB8875 CRC64;

Query Match 67.7%; Score 155; DB 1; Length 424;

Best Local Similarity 100.0%; Pred. No. 2e-11;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 TRSMVFARHLREVGDEFRSHLNSTDAD 43

Db 236 TRSMVFARHLREVGDEFRSHLNSTDAD 265

RESULT 2

Q724NO

ID Q724NO PRELIMINARY; PRT; 429 AA.

AC Q724NO

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)

DE Protein-O-fucosyltransferase (C2Ior180 protein, isoform C).

GN Name=FUT13; Synonyms=POPUT2;

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Martinez-Dunker I., Oriol R., Mollicone R., Candellier J.J., Breton C.,

RT "A new superfamily of protein-O-fucosyltransferases, alpha2-

RT fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and

RT identification of conserved peptide motifs.";

RL Glycobiology 13:1C-5C(2003).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Hapkins R.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko M., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,

RA Krzywinski M.I., Skalska J., Schmutz D.E., Schnerch A., Schein J.E.,

RA Jones S.J., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RA Strausberg R.;

RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY575591; CAE01472.1; -.

DR EMBL; BC064623; AAH64623.1; -.

DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . . ; IEA.

SW Glycosyltransferase; Transferase.

KQ SEQUENCE 429 AA; 49975 MW; 36A4213D905AFFD1 CRC64;

Query Match 67.7%; Score 155; DB 2; Length 429;

Best Local Similarity 100.0%; Pred. No. 2e-11;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 TRSMVFARHLREVGDEFRSHLNSTDAD 43

Db 236 TRSMVFARHLREVGDEFRSHLNSTDAD 265

RESULT 3

Q6EV56

ID Q6EV56 PRELIMINARY; PRT; 429 AA.

AC Q6EV56

DT 01-OCT-2004 (TrEMBLrel. 28, Created)

DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)

DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)

DE Protein-O-fucosyltransferase 2.

GN Name=fut13;

OS Pan troglodytes (Chimpanzee).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

OX NCBI_TaxID=9598;

RN [1]

RP SEQUENCE FROM N.A.

EX PubMed=12966037;

RA Martinez-Dunker I., Mollicone R., Candellier J.J., Breton C.,

RA Oriol R.;

RT "A new superfamily of protein-O-fucosyltransferases, alpha2-

RT fucosyltransferases and alpha6-fucosyltransferases: phylogeny and

RT identification of conserved peptide motifs.";

RL Glycobiology 13:1C-5C(2003).

RN [2]

RP SEQUENCE FROM N.A.

RA Martinez-Dunker I., Oriol R., Mollicone R.;

RT Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ781756; CAH03731.1; --
 KW Glycosyltransferase; Transferrase.
 SQ SEQUENCE 429 AA; 49846 MW; 77A3635BD4698983 CRC64;
 Query Match 67.7%; Score 155; DB 2; Length 429;
 Best Local Similarity 100.0%; Pred. No. 2e-11;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 14 TRSMVFARHLREVGDFRSRHLNSTDDAD 43
 DB 236 TRSMVFARHLREVGDFRSRHLNSTDDAD 265
 RESULT 4
 AAL47681 PRELIMINARY; PRT; 429 AA.
 ID AAL47681;
 AC AAL47681;
 DT 02-MAR-2004 (TrEMBLrel. 27, Created)
 DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
 DE 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
 DE C21orf80 protein form C.
 GN C21orf80.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primata; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Menzel O., Raymond A., Antonarakis S.E., Guipponi M.;
 RT "Disruption and overexpression of the Caenorhabditis elegans pad-2
 RT gene revealed possible roles for C21orf80, its human ortholog, in the
 RT pathogenesis of Down syndrome.";
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY066015; AAL47681.2; -- 36A4213D90SAFFD1 CRC64;
 SQ SEQUENCE 429 AA; 49975 MW; 36A4213D90SAFFD1 CRC64;
 Query Match 67.7%; Score 155; DB 2; Length 429;
 Best Local Similarity 100.0%; Pred. No. 2e-11;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 14 TRSMVFARHLREVGDFRSRHLNSTDDAD 43
 DB 236 TRSMVFARHLREVGDFRSRHLNSTDDAD 265
 RESULT 5
 AAH64623 PRELIMINARY; PRT; 429 AA.
 ID AAH64623
 AC AAH64623
 DT 02-MAR-2004 (TrEMBLrel. 27, Created)
 DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
 DE 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
 DE C21orf80 protein, isoform C.
 GN C21orf80.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RC TISSUE=Brain;
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC011044; AAH11044.1; -- 36A4213D90SAFFD1 CRC64;
 SQ SEQUENCE 141 AA; 16478 MW; 36A4213D90SAFFD1 CRC64;
 Query Match 59.4%; Score 136; DB 2; Length 141;
 Best Local Similarity 100.0%; Pred. No. 1.7e-09;

RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RC TISSUE=Brain;
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC011044; AAH11044.1; -- 36A4213D90SAFFD1 CRC64;
 SQ SEQUENCE 429 AA; 49975 MW; 36A4213D90SAFFD1 CRC64;
 Query Match 67.7%; Score 155; DB 2; Length 429;
 Best Local Similarity 100.0%; Pred. No. 2e-11;
 Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 14 TRSMVFARHLREVGDFRSRHLNSTDDAD 43
 DB 236 TRSMVFARHLREVGDFRSRHLNSTDDAD 265
 RESULT 6
 Q6PUV1 PRELIMINARY; PRT; 141 AA.
 ID Q6PUV1
 AC Q6PUV1;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE POFUT2 protein.
 GN Name=POFUT2; (Human).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RC TISSUE=Brain;
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC011044; AAH11044.1; -- 36A4213D90SAFFD1 CRC64;
 SQ SEQUENCE 141 AA; 16478 MW; 36A4213D90SAFFD1 CRC64;
 Query Match 59.4%; Score 136; DB 2; Length 141;
 Best Local Similarity 100.0%; Pred. No. 1.7e-09;

```

Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 18 MVFARHLREVGDEFRSHLNSTDDAD 43
DB 1 MVFARHLREVGDEFRSHLNSTDDAD 26

RESULT 7
AAH11044
ID AAH11044 PRELIMINARY; PRT; 141 AA.
AC AAH11044;
DT 14-APR-2004 (TrEMBLrel. 27, Created)
DT 14-APR-2004 (TrEMBLrel. 27, Last sequence update)
DT 14-APR-2004 (TrEMBLrel. 27, Last annotation update)
DE C21orf80 protein.
GN C21orf80.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Shenmen C.M., Schuler G.D.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Krzywinski M.I., Skalska U., Smallos D.E.,
RA Jones S.J., Marra M.A.,
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 1-119 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=22354883; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito K., Suzuki H., Yamanaka I., Gotojori T.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gotojori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sadelain A., Schneider C., Sempole C.A., Setou M., Shimada K.,
RA Sullivan R., Takeraka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wysshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume M., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
CC -!- FUNCTION: Catalyzes the reaction that attaches fucose through an
CC O-glycosidic linkage to a conserved serine or threonine residue in
CC EGF domains (by similarity).
CC -!- CATALYTIC ACTIVITY: Transfers an alpha-L-fucosyl residue from GDP-
CC beta-L-fucose to the serine hydroxy group of a protein acceptor.
CC -!- PATHWAY: GLYCOSYLATION.
CC -!- ALTERNATIVE PRODUCTS.
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=A;

```

```
CC CC      IsoId=Q8VH13-1; Sequence=Displayed;
CC CC      Name=2;
CC CC      IsoId=Q8VH13-2; Sequence=VSP_003835;
CC CC      Note=No experimental confirmation available;
CC CC      -!- SIMILARITY: Belongs to the glycosyltransferase family 68.
CC CC
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC
CC CC      EMBL; AF455270; AAL65192.1; -.
CC CC      EMBL; BC018194; -. NOT ANNOTATED_CDS.
CC CC      EMBL; AK009301; BAB26202.1; -.
CC CC      MGD; MGI:1933140; BC003494.
CC CC      Alternative splicing; Fucose metabolism; Glycoprotein;
CC CC      Glycosyltransferase; Signal; Potential.
CC CC      FT SIGNAL 1 21
CC CC      FT CHAIN 22 429
CC CC      FT GDP-fucose protein O-fucosyltransferase
CC CC      FT
CC CC      FT CARBOHYD 189 189 N-linked (GlcNAc...) (Potential).
CC CC      FT CARBOHYD 209 209 N-linked (GlcNAc...) (Potential).
CC CC      FT CARBOHYD 259 259 N-linked (GlcNAc...) (Potential).
CC CC      FT VARSPLIC 382 429 IGVSVTFSPRIHEERILGLDKTYNFCGQEKACEQP
CC CC      FT THWRIAY -> YWHLCPHIFLSDS (in isoform 2).
CC CC      FT /FTID=VSP_003835.
CC CC      SQ SEQUENCE 429 AA; 49429 MW; 0E7AFF5F1CD33560 CRC64;
CC CC
CC CC      Query Match 59.0%; Score 135; DB 1; Length 429;
CC CC      Best Local Similarity 83.9%; Pred. No. 7.7e-09;
CC CC      Matches 26; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
CC CC
CC CC      Qy 14 TRRSVVFARHLREVGDFRSRHLNSTDDAD 44
CC CC      Db 236 TRRSVVFARHLRAVGDFRSRHLNSTDAADK 266
CC CC
CC CC      RESULT 9
CC CC      Q7YRR5 PRELIMINARY; PRT; 429 AA.
CC CC      ID Q7YRR5;
CC CC      AC Q7YRR5;
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Created)
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Last sequence update)
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Last annotation update)
CC CC      DE Hypothetical protein fut13.
CC CC      GN Name=fut13;
CC CC      OS Bos taurus (Bovine).
CC CC      OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC      OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
CC CC      OC Bovinae; Bos.
CC CC      OX NCBI_TaxID=9913;
CC CC      RN [1]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RA Martinez-Dunker I., Mollicone R., Candelier J.J., Breton C., Oriol R.;
CC CC      RT "A new superfamily of protein-O-fucosyltransferases, alpha2-
CC CC      fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and
CC CC      RT identification of conserved peptide motifs.";
CC CC      RL Glycobiology 13:1c-5c(2003).
CC CC      DR EMBL; AJ575655; CAE02608.1; -.
CC CC      KW Hypothetical protein.
CC CC      FT NON TER 1
CC CC      SQ SEQUENCE 429 AA; 49757 MW; 8027B56BDEB02DA6 CRC64;
CC CC
CC CC      Query Match 58.5%; Score 134; DB 2; Length 429;
CC CC      Best Local Similarity 90.0%; Pred. No. 1e-08;
CC CC      Matches 27; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
CC CC
CC CC      Qy 14 TRRSVVFARHLREVGDFRSRHLNSTDDAD 43
CC CC      Db 236 TRRSVVFARHLRAVGDFRSRHLNSTDVAD 265
CC CC
CC CC      RESULT 10
CC CC      Q7YRR5 PRELIMINARY; PRT; 459 AA.
CC CC      ID Q7YRR5;
CC CC      AC Q7YRR5;
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Created)
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Last sequence update)
CC CC      DT 01-WAR-2004 (TrEMBLrel. 26; Last annotation update)
CC CC      DE Protein O-fucosyltransferase 2.
CC CC      GN Name=pofut2;
CC CC      OS Bos taurus (Bovine).
CC CC      OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC      OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
CC CC      OC Bovinae; Bos.
CC CC      OX NCBI_TaxID=9913;
CC CC      RN [1]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RA Lorient C., Germot A., Dupuy F., Maftah A.;
CC CC      RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
CC CC      DR EMBL; AY344582; AAQ02334.1; -.
CC CC      DR GO; GO:001675; F:transferase activity, transferring glycosyl...; IEA.
CC CC      KW Glycosyltransferase; Transferase.
CC CC      SQ SEQUENCE 459 AA; 52811 MW; 207CEFFE7D3CF7E5 CRC64;
CC CC
CC CC      Query Match 58.5%; Score 134; DB 2; Length 459;
CC CC      Best Local Similarity 90.0%; Pred. No. 1.1e-08;
CC CC      Matches 27; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
CC CC
CC CC      Qy 14 TRRSVVFARHLREVGDFRSRHLNSTDDAD 43
CC CC      Db 236 TRRSVVFARHLRAVGDFRSRHLNSTDVAD 265
CC CC
CC CC      RESULT 11
CC CC      Q7TLN6 PRELIMINARY; PRT; 380 AA.
CC CC      ID Q7TLN6;
CC CC      AC Q7TLN6;
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Created)
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Last sequence update)
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Last annotation update)
CC CC      DE Hypothetical protein FUT13 (Fragment).
CC CC      GN Name=FUT13;
CC CC      OS Gallus gallus (Chicken).
CC CC      OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC      OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
CC CC      OC Gallus.
CC CC      OX NCBI_TaxID=9031;
CC CC      RN [1]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RA Martinez-Dunker I., Mollicone R., Candelier J.J., Breton C., Oriol R.;
CC CC      RT "A new superfamily of protein-O-fucosyltransferases, alpha2-
CC CC      fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and
CC CC      RT identification of conserved peptide motifs.";
CC CC      RL Glycobiology 13:1c-5c(2003).
CC CC      DR EMBL; AJ575657; CAE02610.1; -.
CC CC      KW Hypothetical protein.
CC CC      FT NON TER 1
CC CC      SQ SEQUENCE 380 AA; 44572 MW; F697E4600621291B CRC64;
CC CC
CC CC      Query Match 52.0%; Score 119; DB 2; Length 380;
CC CC      Best Local Similarity 73.3%; Pred. No. 7.8e-07;
CC CC      Matches 22; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
CC CC
CC CC      Qy 14 TRRSVVFARHLREVGDFRSRHLNSTDDAD 43
CC CC      Db 187 TRRSVVFARHLRAVGDFRSRHLNSTDEAD 216
CC CC
CC CC      RESULT 12
CC CC      Q6EV53 PRELIMINARY; PRT; 386 AA.
CC CC      ID Q6EV53;
CC CC      AC Q6EV53;
CC CC      DT 01-OCT-2004 (TrEMBLrel. 28; Created)
```

```
CC CC      IsoId=Q8VH13-1; Sequence=Displayed;
CC CC      Name=2;
CC CC      IsoId=Q8VH13-2; Sequence=VSP_003835;
CC CC      Note=No experimental confirmation available;
CC CC      -!- SIMILARITY: Belongs to the glycosyltransferase family 68.
CC CC
CC CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC      the European Bioinformatics Institute. There are no restrictions on its
CC CC      use by non-profit institutions as long as its content is in no way
CC CC      modified and this statement is not removed. Usage by and for commercial
CC CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC      or send an email to license@isb-sib.ch).
CC CC
CC CC      EMBL; AF455270; AAL65192.1; -.
CC CC      EMBL; BC018194; -. NOT ANNOTATED_CDS.
CC CC      EMBL; AK009301; BAB26202.1; -.
CC CC      MGD; MGI:1933140; BC003494.
CC CC      Alternative splicing; Fucose metabolism; Glycoprotein;
CC CC      Glycosyltransferase; Signal; Potential.
CC CC      FT SIGNAL 1 21
CC CC      FT CHAIN 22 429
CC CC      FT GDP-fucose protein O-fucosyltransferase
CC CC      FT
CC CC      FT CARBOHYD 189 189 N-linked (GlcNAc...) (Potential).
CC CC      FT CARBOHYD 209 209 N-linked (GlcNAc...) (Potential).
CC CC      FT CARBOHYD 259 259 N-linked (GlcNAc...) (Potential).
CC CC      FT VARSPLIC 382 429 IGVSVTFSPRIHEERILGLDKTYNFCGQEKACEQP
CC CC      FT THWRIAY -> YWHLCPHIFLSDS (in isoform 2).
CC CC      FT /FTID=VSP_003835.
CC CC      SQ SEQUENCE 429 AA; 49429 MW; 0E7AFF5F1CD33560 CRC64;
CC CC
CC CC      Query Match 59.0%; Score 135; DB 1; Length 429;
CC CC      Best Local Similarity 83.9%; Pred. No. 7.7e-09;
CC CC      Matches 26; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
CC CC
CC CC      Qy 14 TRRSVVFARHLREVGDFRSRHLNSTDDAD 44
CC CC      Db 236 TRRSVVFARHLRAVGDFRSRHLNSTDAADK 266
CC CC
CC CC      RESULT 9
CC CC      Q7YRR5 PRELIMINARY; PRT; 429 AA.
CC CC      ID Q7YRR5;
CC CC      AC Q7YRR5;
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Created)
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Last sequence update)
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Last annotation update)
CC CC      DE Hypothetical protein fut13.
CC CC      GN Name=fut13;
CC CC      OS Bos taurus (Bovine).
CC CC      OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC      OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
CC CC      OC Bovinae; Bos.
CC CC      OX NCBI_TaxID=9913;
CC CC      RN [1]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RA Martinez-Dunker I., Mollicone R., Candelier J.J., Breton C., Oriol R.;
CC CC      RT "A new superfamily of protein-O-fucosyltransferases, alpha2-
CC CC      fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and
CC CC      RT identification of conserved peptide motifs.";
CC CC      RL Glycobiology 13:1c-5c(2003).
CC CC      DR EMBL; AJ575655; CAE02608.1; -.
CC CC      KW Hypothetical protein.
CC CC      FT NON TER 1
CC CC      SQ SEQUENCE 429 AA; 49757 MW; 8027B56BDEB02DA6 CRC64;
CC CC
CC CC      Query Match 58.5%; Score 134; DB 2; Length 429;
CC CC      Best Local Similarity 90.0%; Pred. No. 1e-08;
CC CC      Matches 27; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
CC CC
CC CC      Qy 14 TRRSVVFARHLREVGDFRSRHLNSTDDAD 43
CC CC      Db 236 TRRSVVFARHLRAVGDFRSRHLNSTDVAD 265
CC CC
CC CC      RESULT 10
CC CC      Q7YRR5 PRELIMINARY; PRT; 459 AA.
CC CC      ID Q7YRR5;
CC CC      AC Q7YRR5;
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Created)
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Last sequence update)
CC CC      DT 01-WAR-2004 (TrEMBLrel. 26; Last annotation update)
CC CC      DE Protein O-fucosyltransferase 2.
CC CC      GN Name=pofut2;
CC CC      OS Bos taurus (Bovine).
CC CC      OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC      OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
CC CC      OC Bovinae; Bos.
CC CC      OX NCBI_TaxID=9913;
CC CC      RN [1]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RA Lorient C., Germot A., Dupuy F., Maftah A.;
CC CC      RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
CC CC      DR EMBL; AY344582; AAQ02334.1; -.
CC CC      DR GO; GO:001675; F:transferase activity, transferring glycosyl...; IEA.
CC CC      KW Glycosyltransferase; Transferase.
CC CC      SQ SEQUENCE 459 AA; 52811 MW; 207CEFFE7D3CF7E5 CRC64;
CC CC
CC CC      Query Match 58.5%; Score 134; DB 2; Length 459;
CC CC      Best Local Similarity 90.0%; Pred. No. 1.1e-08;
CC CC      Matches 27; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
CC CC
CC CC      Qy 14 TRRSVVFARHLREVGDFRSRHLNSTDDAD 43
CC CC      Db 236 TRRSVVFARHLRAVGDFRSRHLNSTDVAD 265
CC CC
CC CC      RESULT 11
CC CC      Q7TLN6 PRELIMINARY; PRT; 380 AA.
CC CC      ID Q7TLN6;
CC CC      AC Q7TLN6;
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Created)
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Last sequence update)
CC CC      DT 01-OCT-2003 (TrEMBLrel. 25; Last annotation update)
CC CC      DE Hypothetical protein FUT13 (Fragment).
CC CC      GN Name=FUT13;
CC CC      OS Gallus gallus (Chicken).
CC CC      OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC      OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
CC CC      OC Gallus.
CC CC      OX NCBI_TaxID=9031;
CC CC      RN [1]
CC CC      RP SEQUENCE FROM N.A.
CC CC      RA Martinez-Dunker I., Mollicone R., Candelier J.J., Breton C., Oriol R.;
CC CC      RT "A new superfamily of protein-O-fucosyltransferases, alpha2-
CC CC      fucosyltransferases, and alpha6-fucosyltransferases: phylogeny and
CC CC      RT identification of conserved peptide motifs.";
CC CC      RL Glycobiology 13:1c-5c(2003).
CC CC      DR EMBL; AJ575657; CAE02610.1; -.
CC CC      KW Hypothetical protein.
CC CC      FT NON TER 1
CC CC      SQ SEQUENCE 380 AA; 44572 MW; F697E4600621291B CRC64;
CC CC
CC CC      Query Match 52.0%; Score 119; DB 2; Length 380;
CC CC      Best Local Similarity 73.3%; Pred. No. 7.8e-07;
CC CC      Matches 22; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
CC CC
CC CC      Qy 14 TRRSVVFARHLREVGDFRSRHLNSTDDAD 43
CC CC      Db 187 TRRSVVFARHLRAVGDFRSRHLNSTDEAD 216
CC CC
CC CC      RESULT 12
CC CC      Q6EV53 PRELIMINARY; PRT; 386 AA.
CC CC      ID Q6EV53;
CC CC      AC Q6EV53;
CC CC      DT 01-OCT-2004 (TrEMBLrel. 28; Created)
```

DT 01-OCT-2004 (TReMBLrel. 28, Last sequence update)
 DE 01-OCT-2004 (TReMBLrel. 28, Last annotation update)
 DE Protein-O-fucosyltransferase 2 (Fragment).
 GN Name=fut13;
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
 OC Tetraodontidae; Tetraodontidae; Takifugu.
 OX NCBI_TaxID=31033;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=12966037;
 RA Martinez-Duncker I., Mollicone R., Candellier J.J., Breton C.,
 Oriol R.;
 RA "A new superfamily of protein-O-fucosyltransferases, alpha2-
 RT fucosyltransferases and alpha3-fucosyltransferases: phylogeny and
 RT identification of conserved peptide motifs.";
 RL Glycobiology 13:1C-5C(2003).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Martinez-Duncker I., Oriol R., Mollicone R.;
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ781759; CAH03734.1; -;
 KW Glycosyltransferase; Transferase.
 FT NON TER 1
 SQ SEQUENCE 386 AA; 44953 MW; CA80543FC5DB7A1D CRC64;
 Query Match 51.1%; Score 117; DB 2; Length 386;
 Best Local Similarity 73.3%; Pred. No. 1.4e-06;
 Matches 22; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 14 TRSMVFARHLREVGDEFRSHLNSTDAD 43
 Db |||||:||||:||||:||||:||||:|
 193 TRSMVFARHLRLIGDNFRATHLNSTDISD 222
 RESULT 13
 Q99LA7 PRELIMINARY; PRT; 201 AA.
 AC Q99LA7;
 DT 01-JUN-2001 (TReMBLrel. 17, Created)
 DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Pofut2 protein.
 GN Name=pofut2;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 Brownstein M.J., Uscin T.B., Toshiyuki S., Carninci P., Prange C.,
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guarnatone P.H.,
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 Krzywinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
 Jones S.J., Marra M.A.;
 RP "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mix FVB/N;
 RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
 RA Strausberg R.;
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC003494; AAH03494.1; -;
 DR MGD; MGI:1933140; BC003494.
 SQ SEQUENCE 201 AA; 22901 MW; 66889A53389AD124 CRC64;
 Query Match 34.5%; Score 79; DB 2; Length 201;
 Best Local Similarity 83.3%; Pred. No. 0.056;
 Matches 15; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 27 VGDEFRSHLNSTDAD 44
 Db |||||:||||:||||:||||:||||:|
 21 VGDEFRSHLNSTDADK 38
 RESULT 14
 Q7YZS7 PRELIMINARY; PRT; 434 AA.
 AC Q7YZS7;
 DT 01-OCT-2003 (TReMBLrel. 25, Created)
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Hypothetical protein fut13.
 GN Name=fut13;
 OS Clona intestinalis.
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 OC Phlebobranchia; Clonidae; Clona.
 OX NCBI_TaxID=7719;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Martinez-Duncker I., Mollicone R., Candellier J.J., Breton C., Oriol R.;
 RT "A new superfamily of protein-O-fucosyltransferases, alpha2-
 RT fucosyltransferases, and alpha3-fucosyltransferases: phylogeny and
 RT identification of conserved peptide motifs.";
 RL Glycobiology 13:1C-5C(2003).
 DR EMBL; AJ575656; CAE02609.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 434 AA; 50605 MW; D853C746DEB896C CRC64;
 Query Match 32.8%; Score 75; DB 2; Length 434;
 Best Local Similarity 53.3%; Pred. No. 0.42;
 Matches 16; Conservative 6; Mismatches 8; Indels 0; Gaps 0;
 QY 15 RSMVFARHLREVGDEFRSHLNSTDAD 44
 Db |||||:||||:||||:||||:||||:|
 234 RSMVFADALIKTADQFRSKYLDSDVDK 263
 RESULT 15
 Q7K729 PRELIMINARY; PRT; 423 AA.
 AC Q7K729;
 DT 05-JUL-2004 (TReMBLrel. 27, Created)
 DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
 DE EG-BACN32G11.6 protein.
 GN Name=EG-BACN32G11.6;
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Mottier, Cadieu, Dreano, Lelaure, Galibert F.;
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
RA Berios P.; (AUG-1999) to the EMBL/GenBank/DBJ databases.
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL035632; CAB53644.1; -.
DR InterPro; IPR001412; tRNA-synt 1.
DR PROSITE; PS00178; AA tRNA LIGASE I; UNKNOWN 1.
SQ SEQUENCE 423 AA; 48995 MW; 28DD7D64E6B39015 CRC64;

Query Match 32.3%; Score 74; DB 2; Length 423;
Best Local Similarity 53.3%; Pred.No. 0.56;
Matches 16; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 11 PPOTRRSMVFARHLREYGVDBFRGRHLNSTD 40
Db ||||| ||||| :||| :|||
211 FWQARRSRFARRLEQVAADFRRQALDTTD 240
||| ||| :||| :|||

Search completed: November 10, 2004, 13:38:25
Job time : 54.7271 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:32:37 ; Search time 15.544 Seconds
(without alignments)
191.991 Million cell updates/sec

Title: US-10-092-750-30
Perfect score: 219
Sequence: 1 GLKLTVAASMDRVPKVTSP.....IARENEPERLGLNGIAETT 45

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64.5	29.5	1693	3 US-08-840-316-1	Sequence 1, Appli
2	64.5	29.5	1693	3 US-08-809-523-1	Sequence 1, Appli
3	64.5	29.5	1693	3 US-08-471-971-1	Sequence 1, Appli
4	64.5	29.5	1693	4 US-09-402-776-1	Sequence 1, Appli
5	64.5	29.5	1693	4 US-08-470-246-1	Sequence 1, Appli
6	64.5	29.5	1693	4 US-08-316-785-1	Sequence 1, Appli
7	64.5	29.5	1693	4 US-09-724-475-1	Sequence 1, Appli
8	64.5	29.5	1693	5 PCT-US93-08849A-1	Sequence 1, Appli
9	64.5	29.5	1693	5 PCT-US93-08849-1	Sequence 1, Appli
10	62	28.3	153	4 US-09-270-767-35615	Sequence 35615, A
11	62	28.3	153	4 US-09-270-767-50832	Sequence 50832, A
12	61.5	28.1	1427	4 US-09-538-092-1044	Sequence 1044, Ap
13	55.5	25.3	356	4 US-09-252-991A-27056	Sequence 27056, A
14	55.5	25.3	521	4 US-09-724-797-62	Sequence 62, Appl
15	54.5	24.9	1693	3 US-08-478-507-7	Sequence 7, Appli
16	54.5	24.9	1693	3 US-09-128-275A-7	Sequence 7, Appli
17	54.5	24.9	1693	3 US-09-553-427-7	Sequence 7, Appli
18	54	24.7	442	4 US-09-252-991A-30607	Sequence 30607, A
19	54	24.7	484	4 US-09-602-787A-26	Sequence 26, Appl
20	54	24.7	491	3 US-09-362-899-3	Sequence 3, Appli
21	52	23.7	170	4 US-09-252-991A-18663	Sequence 18663, A
22	52	23.7	282	3 US-09-222-939-17	Sequence 17, Appl
23	52	23.7	282	4 US-10-023-528-17	Sequence 17, Appl
24	52	23.7	332	4 US-09-134-000C-4723	Sequence 4723, Ap
25	51	23.3	403	4 US-09-252-991A-25876	Sequence 25876, A
26	51	23.3	449	4 US-10-140-002-224	Sequence 224, App
27	50.5	23.1	460	4 US-09-252-991A-32153	Sequence 32153, A

28	50.5	23.1	575	3 US-09-171-461-30	Sequence 30, Appl
29	50.5	23.1	575	4 US-09-970-711-30	Sequence 30, Appl
30	50	22.8	281	4 US-09-252-991A-18840	Sequence 18840, A
31	50	22.8	311	4 US-09-248-766A-16178	Sequence 16178, A
32	50	22.8	314	4 US-09-270-767-42409	Sequence 42409, A
33	50	22.8	449	4 US-09-489-039A-8009	Sequence 8009, Ap
34	50	22.8	489	4 US-08-190-204-2	Sequence 2, Appli
35	50	22.8	489	5 PCT-US93-11110-1	Sequence 1, Appli
36	50	22.8	489	6 5221789-1	Patent No. 5221789
37	50	22.8	695	4 US-09-270-767-43792	Sequence 43792, A
38	50	22.8	836	4 US-09-538-092-173	Sequence 173, App
39	49.5	22.6	1665	4 US-09-543-681A-4476	Sequence 4476, App
40	49	22.4	152	3 US-08-679-493A-203	Sequence 203, App
41	49	22.4	161	4 US-09-270-767-56844	Sequence 56844, A
42	49	22.4	170	4 US-09-270-767-36791	Sequence 36791, A
43	49	22.4	170	4 US-09-270-767-52008	Sequence 52008, A
44	49	22.4	252	4 US-09-270-767-47090	Sequence 47090, A
45	49	22.4	286	4 US-09-328-352-6705	Sequence 6705, Ap

ALIGNMENTS

RESULT 1
US-08-840-316-1
; Sequence 1, Application US/08940316
; Patent No. 6054567
; GENERAL INFORMATION:
; APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,
; TITLE OF INVENTION: Recombinant Proteins Of
; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
; TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/840,316
; FILING DATE: 11-APR-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Richard W. Bork
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4255
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1693 AMINO ACID RESIDUES
; TYPE: AMINO ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
; US-08-840-316-1

Query Match 29.5%; Score 64.5; DB 3; Length 1693;
Best Local Similarity 37.3%; Pred. No. 1.6;
Matches 19; Conservative 7; Mismatches 14; Indels 11; Gaps 2;

QY 1 GLKLTVAASMDR-----VPKVTSSA-----ISSIARENHEPERLGLG 40
 Db 616 GLEVRYAAGLDHRAVFAPGVSFSPAGVTAFCALYRFNREARLSLTG 666

RESULT 2

US-08-803-523-1
 ; Sequence 1, Application US/08809523
 ; Patent No. 6207416
 ; GENERAL INFORMATION:
 ; APPLICANT: Tearev, Sergei. A., Emerson,
 ; APPLICANT: Suzanne U., Purcell, Robert H.,
 ; TITLE OF INVENTION: Recombinant Proteins Of
 ; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis B And Their
 ; TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
 ; NUMBER OF SEQUENCES: 107
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/803/523
 ; FILING DATE: 28-MAY-1997
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/13102
 ; FILING DATE: 03-OCT-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US08/316,765
 ; FILING DATE: 03-OCT-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 07/947,263
 ; FILING DATE: 18-SEP-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Richard W. Bork
 ; REGISTRATION NUMBER: 36,459
 ; REFERENCE/DOCKET NUMBER: 2026-4032US4
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1693 AMINO ACID RESIDUES
 ; TYPE: AMINO ACID
 ; STRANDEDNESS: UNKNOWN
 ; TOPOLOGY: UNKNOWN
 ; US-08-803-523-1

Query Match 29.5%; Score 64.5; DB 3; Length 1693;
 Best Local Similarity 37.3%; Pred. No. 1.6;
 Matches 19; Conservative 7; Mismatches 14; Indels 11; Gaps 2;

QY 1 GLKLTVAASMDR-----VPKVTSSA-----ISSIARENHEPERLGLG 40
 Db 616 GLEVRYAAGLDHRAVFAPGVSFSPAGVTAFCALYRFNREARLSLTG 666

RESULT 3

US-08-471-971-1
 ; Sequence 1, Application US/08471971
 ; Patent No. 6287759
 ; GENERAL INFORMATION:
 ; APPLICANT: Tearev, Sergei. A., Emerson,
 ; APPLICANT: Suzanne U., Purcell, Robert H.,
 ; TITLE OF INVENTION: Recombinant Proteins Of
 ; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis B And Their

; TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
 ; NUMBER OF SEQUENCES: 107
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/471,971
 ; FILING DATE: 06-JUN-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US08/316,765
 ; FILING DATE: 03-OCT-1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US07/947,263
 ; FILING DATE: 18-SEP-1992
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Richard W. Bork
 ; REGISTRATION NUMBER: 36,459
 ; REFERENCE/DOCKET NUMBER: 2026-4032US2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 758-4800
 ; TELEFAX: (212) 751-6849
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1693 AMINO ACID RESIDUES
 ; TYPE: AMINO ACID
 ; STRANDEDNESS: UNKNOWN
 ; TOPOLOGY: UNKNOWN
 ; US-08-471-971-1

QY 1 GLKLTVAASMDR-----VPKVTSSA-----ISSIARENHEPERLGLG 40
 Db 616 GLEVRYAAGLDHRAVFAPGVSFSPAGVTAFCALYRFNREARLSLTG 666

RESULT 4
 US-09-402-776-1
 ; Sequence 1, Application US/09402776
 ; Patent No. 6458562
 ; GENERAL INFORMATION:
 ; APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,
 ; APPLICANT: Tearev, Sergei. A., and Robinson, Robin A.
 ; TITLE OF INVENTION: Recombinant Proteins Of
 ; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis B And Their
 ; TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
 ; NUMBER OF SEQUENCES: 111
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1

Query Match 29.5%; Score 64.5; DB 3; Length 1693;
 Best Local Similarity 37.3%; Pred. No. 1.6;
 Matches 19; Conservative 7; Mismatches 14; Indels 11; Gaps 2;


```

;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/402,776
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/840,316
; FILING DATE: 11-APR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Richard W. Bork
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4255
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1693 AMINO ACID RESIDUES
; TYPE: AMINO ACID
; STRANDEDNESS: UNKNOWN
; TOPOLOGY: UNKNOWN
; US-09-402-776-1
;
; Query Match 29.5%; Score 64.5; DB 4; Length 1693;
; Best Local Similarity 37.3%; Pred.No.1.6;
; Matches 19; Conservative 7; Mismatches 14; Indels 11; Gaps 2;
;
; QY 1 GLKLATVAASMDR---VPKVTSSA-----ISSIARENHHEPERLGLNG 40
; DB 616 GLEVRVYAAGLDHRAVAPGVSPRSAPCEVTAFCASLYRFNREARUSLTG 666
;
; RESULT 5
; US-08-470-246-1
; Sequence 1, Application US/08470246
; Patent No. 696242
; GENERAL INFORMATION:
; APPLICANT: Tsarev, Sergei. A., Emerson,
; APPLICANT: Suzanne U., Purcell, Robert H.
; TITLE OF INVENTION: Recombinant Proteins Of
; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
; TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,246
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US08/316,765
; FILING DATE: 03-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US07/947,263
; FILING DATE: 18-SEP-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Richard W. Bork
; REGISTRATION NUMBER: 36,459
; REFERENCE/DOCKET NUMBER: 2026-4032US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
;

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08849A
FILING DATE: 17-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US07/947,263
FILING DATE: 18-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: William S. Feiler
REGISTRATION NUMBER: 26,728
REFERENCE/DOCKET NUMBER: 2026-4032 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 AMINO ACID RESIDUES
TYPE: AMINO ACID
STRANDEDNESS: UNKNOWN
TOPOLOGY: UNKNOWN
PCT-US93-08849A-1

Query Match 29.5%; Score 64.5; DB 5; Length 1693;
Best Local Similarity 37.3%; Pred. No. 1.6;
Matches 19; Conservative 7; Mismatches 14; Indels 11;

QY 1 GIKIATVAASMDR----VPKVTSSA-----ISSIATRNHEPERGLGLNG 40
DB 616 GLEVRVYVAAGLDHRVAFPGVSPRSARPEVTAFCSALYRFNREARQLSLTG 666

RESULT 9
PCT-US93-08849-1
Sequence 1, Application PC/TUS9308849
GENERAL INFORMATION:
APPLICANT: Tsarev, Sergei A., Emerson,
APPLICANT: Suzanne U., Purcell, Robert H.
TITLE OF INVENTION: Recombinant Proteins Of
TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESS: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08849
FILING DATE: 17-SEP-1993
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/947,263
FILING DATE: 18-SEP-1992
NAME:
ATTORNEY/AGENT INFORMATION:
NAME: Bork, Richard, W.
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 1:

```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1693 amino acid residues
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
PCT-US93-08849-1

Query Match      29.5%; Score 64.5; DB 5; Length 1693;
Best Local Similarity 37.3%; Pred. No. 1.6;
Matches 19; Conservative 7; Mismatches 14; Indels 11; Gaps 2;

QY 1 GLKLATVAASMDRVPKVPSSAIISSIAIRENHEPERLGLNG 40
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 616 GLEVRVVAAGLDRHRAVPAGVSPASPCVETAFCSALYRFRNRAQRLSLTG 666

RESULT 10
US-09-270-767-35615
; Sequence 35615, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 35615
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-35615

Query Match      28.3%; Score 62; DB 4; Length 153;
Best Local Similarity 35.2%; Pred. No. 0.15;
Matches 19; Conservative 5; Mismatches 14; Indels 16; Gaps 2;

QY 1 GLKLATVAASMDRVPKVPSSAIISSIAIRENHEPERLGLNG 40
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 75 GLRIA-EATPTPTTPTSEATPSSRSERKTHGSCIDFSTHEPAELGALG 126

RESULT 11
US-09-270-767-50832
; Sequence 50832, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 50832
; LENGTH: 153
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-50832

Query Match      28.3%; Score 62; DB 4; Length 153;
Best Local Similarity 35.2%; Pred. No. 0.15;
Matches 19; Conservative 5; Mismatches 14; Indels 16; Gaps 2;

QY 1 GLKLATVAASMDRVPKVPSSAIISSIAIRENHEPERLGLNG 40
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 75 GLRIA-EATPTPTTPTSEATPSSRSERKTHGSCIDFSTHEPAELGALG 126

RESULT 12
US-09-538-092-1044
; Sequence 1044, Application US/09538092

; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqformatter Version 0.9
; SEQ ID NO 1044
; LENGTH: 1427
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P30622
US-09-538-092-1044

Query Match      28.1%; Score 61.5; DB 4; Length 1427;
Best Local Similarity 37.2%; Pred. No. 3.7;
Matches 16; Conservative 10; Mismatches 14; Indels 3; Gaps 2;

QY 4 LATVAASMDRVPKVPSSAIISSIAIRENHEPERLGLNGIAETT 45
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 300 MATTSASLKRSPSASSLSMSVSVSSRPSRTGL--LTETS 340

RESULT 13
US-09-252-991A-27056
; Sequence 27056, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27056
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27056

Query Match      25.3%; Score 55.5; DB 4; Length 356;
Best Local Similarity 36.8%; Pred. No. 4.6;
Matches 14; Conservative 9; Mismatches 14; Indels 1; Gaps 1;

QY 7 VAASMDRVPKVP--SSAIISSIAIRENHEPERLGLNGIAE 43
      :|||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 113 LAALVDALHQLVFISSMISLERKGHAPTLLYERGIAE 150

RESULT 14
US-09-724-797-62
; Sequence 62, Application US/09724797
; Patent No. 6733998
; GENERAL INFORMATION:
; APPLICANT: Jon S. THORSON
; TITLE OF INVENTION: MICROMONOSPORA ECHINOSPORA GENES
; TITLE OF INVENTION: ENCODING FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: CALICHEAMICIN AND SELF-RESISTANCE THERETO
```

FILED DATE: 17 JUN 1964
ATTORNEY/AGENT INFORMATION

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 16:36:12 ; Search time 48.342 Seconds
(without alignments)
328.807 Million cell updates/sec

Title: US-10-092-750-30

Perfect score: 219

Sequence: 1 GLKLTVAASMDRVPKVTPTSSAISSIAARENHEPERLGLNGIAETT 45

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	219	100.0	45	14	US-10-092-750-30
2	118	53.9	110	14	US-10-029-386-27800
3	96	43.8	742	14	US-10-104-047-2962
4	65	29.7	2045	16	US-10-437-963-109099
5	63	28.8	488	15	US-10-424-599-221067
6	63	28.8	490	15	US-10-425-114-38453
7	62	28.3	786	16	US-10-437-963-184006
8	60.5	27.6	121	16	US-10-767-701-47653
9	59.5	27.2	354	13	US-10-087-192-120
10	59	26.9	1446	16	US-10-437-963-184040
11	58.5	26.7	1372	17	US-10-425-115-291980
12	58	26.5	208	15	US-10-425-114-38059
13	58	26.5	276	17	US-10-425-115-223595

14	58	26.5	296	15	US-10-282-122A-46968
15	58	26.5	291	15	US-10-425-114-64239
16	58	26.5	26.5	17	US-10-425-115-223592
17	58	26.5	1256	16	US-10-437-963-112766
18	57	26.0	263	16	US-10-437-963-109678
19	56.5	25.8	625	17	US-10-425-115-291983
20	56	25.6	220	16	US-10-437-963-134896
21	56	25.6	285	16	US-10-437-963-107627
22	56	25.6	498	16	US-10-781-014-564
23	56	25.6	606	16	US-10-437-963-132399
24	56	25.6	675	9	US-09-738-626-5917
25	56	25.6	735	16	US-10-437-963-203833
26	56	25.6	1025	16	US-10-437-963-177955
27	56	25.6	1954	16	US-10-437-963-178698
28	56	25.6	1990	16	US-10-437-963-178706
29	55.5	25.3	236	10	US-09-934-455-294
30	55.5	25.3	236	14	US-10-225-066A-692
31	55.5	25.3	236	15	US-10-374-780A-2214
32	55.5	25.3	439	13	US-10-087-192-117
33	55.5	25.3	1169	9	US-09-801-368-106
34	55.5	25.3	1169	14	US-10-369-493-22288
35	55	25.1	253	14	US-10-156-761-14536
36	55	25.1	378	15	US-10-282-122A-56832
37	55	25.1	766	16	US-10-437-963-126177
38	55	25.1	1348	16	US-10-437-963-113840
39	55	25.1	1615	16	US-10-437-963-160697
40	55	25.1	1645	16	US-10-437-963-203480
41	54.5	24.9	362	17	US-10-425-115-341313
42	54.5	24.9	1693	10	US-09-851-410-7
43	54	24.7	397	16	US-10-437-963-143466
44	54	24.7	451	15	US-10-424-599-222880
45	54	24.7	484	15	US-10-627-476-26

ALIGNMENTS

RESULT 1

US-10-092-750-30
; Sequence 30, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-XL
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 45
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-30

Query Match 100.0%; Score 219; DB 14; Length 45;
Best Local Similarity 100.0%; Pred. No. 2.5e-23;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GLKLTVAASMDRVPKVTPTSSAISSIAARENHEPERLGLNGIAETT 45

Db 1 GLKLTVAASMDRVPKVTPTSSAISSIAARENHEPERLGLNGIAETT 45

RESULT 2

US-10-029-386-27800
; Sequence 27800, Application US/10029386
; Publication No. US20030394704A1
; GENERAL INFORMATION:

```
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR Q
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEWICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence-Listing Engine vers. 1.1
; SEQ ID NO 27800
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC015763.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.2
; OTHER INFORMATION: SWISSPROT HIT: P38739, EVALUATE 1.30e-01
US-10-029-386-27800

Query Match 53.9%; Score 118; DB 14; Length 110;
Best Local Similarity 100.0%; Pred. No. 1.5e-08;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LKLATVAASMDRVPKVPKTPSSAIISSIA 27
Db 85 LKLATVAASMDRVPKVPKTPSSAIISSIA 110

RESULT 3
US-10-104-047-2962
; Sequence 2962, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1el full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2962
; LENGTH: 742
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2962

Query Match 43.8%; Score 96; DB 14; Length 742;
Best Local Similarity 52.3%; Pred. No. 0.00021;
Matches 23; Conservative 7; Mismatches 12; Indels 2; Gaps 1;

QY 2 LKLATVAASMDRVPKVPKTPSSAIISSIAIARENHEPERLGLNGIAETT 45
Db 695 LKLA--AAANQVSKVPSSSVSDVPRENHSEKPAIINSIADNT 736

RESULT 4
US-10-437-963-190999
; Sequence 190999, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 190999
; LENGTH: 2045
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(2045)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_8735C.1.pap
US-10-437-963-190999

Query Match 29.7%; Score 65; DB 16; Length 2045;
Best Local Similarity 40.6%; Pred. No. 19;
Matches 13; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 5 ATVAASMDRVPKVPKTPSSAIISSIAIARENHEPERL 36
Db 122 ASIGTSMISEPEKDFSAKPCISNKNHEPTRI 153

RESULT 5
US-10-424-599-221067
; Sequence 221067, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 221067
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_41653C.1.pap
US-10-424-599-221067

Query Match 28.8%; Score 63; DB 15; Length 488;
Best Local Similarity 31.8%; Pred. No. 5.7;
Matches 14; Conservative 12; Mismatches 16; Indels 2; Gaps 1;

QY 3 KLATVAASMDRVPKVPKTPSSAIISSIAIAREN--HEPERLGLNGIAET 44
Db 32 ELCQVAANMERLENVISSTDLSQLASDTTLYDPSNIGLGSWVD 75

RESULT 6
US-10-425-114-38253
; Sequence 38253, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 38253
; LENGTH: 490
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3106-102-C2_FLI.pep
US-10-425-114-38253

Query Match 28.8%; Score 63; DB 15; Length 490;
Best Local Similarity 31.8%; Pred. No. 5.8;
Matches 14; Conservative 12; Mismatches 16; Indels 2; Gaps 1;

QY 3 KLTAAASMDRVPKVPSSAIISSIAAREN--HEPERLNGIAET 44
Db 34 ELCQVAANERLENVISSTDLSQLASDTTLYDPSNIGLSWDT 77

RESULT 7

US-10-437-963-184006
; Sequence 184006, Application US/10437963
; Publication No. US20040123343A1

; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 184006
; LENGTH: 786
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_81040C.1.pep
US-10-437-963-184006

Query Match 28.3%; Score 62; DB 16; Length 786;
Best Local Similarity 37.5%; Pred. No. 15;
Matches 12; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

QY 5 ATVAASMDRVPKVPSSAIISSIAARENHEPERL 36
Db 71 ABIGTSMTEPEKDFSAKTLCSDRNHEPTRI 102

RESULT 8

US-10-767-701-47653
; Sequence 47653, Application US/10767701
; Publication No. US20040172684A1

; GENERAL INFORMATION:
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei

; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 47653

; LENGTH: 121
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3476-038-P1-K1-C3.pep
US-10-767-701-47653

Query Match 27.6%; Score 60.5; DB 16; Length 121;
Best Local Similarity 40.0%; Pred. No. 2.2;
Matches 18; Conservative 7; Mismatches 17; Indels 3; Gaps 2;

QY 2 LKLTAA--SMDRVPKVPSSAIISSIAARENHEPERLNGIAETT 45
Db 33 LAAVVAAMPAPMSRSGATSLARRN--RRIGADALTTT 75

RESULT 9

US-10-087-192-120
; Sequence 120, Application US/10087192
; Publication No. US20020182586A1

; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 529452000122
; CURRENT APPLICATION NUMBER: US/10/087,192
; CURRENT FILING DATE: 2002-03-01
; PRIOR APPLICATION NUMBER: US 09/747,377
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 2059
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 120
; LENGTH: 354
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-087-192-120

Query Match 27.2%; Score 59.5; DB 13; Length 354;
Best Local Similarity 38.9%; Pred. No. 12;
Matches 14; Conservative 8; Mismatches 13; Indels 1; Gaps 1;

QY 4 LATVAASMDRVPKVPSSAIISSIAARE-NHEPERLGL 38
Db 304 MATTSASLAKRSPSASSLSMSVSVSSVSPSTGL 339

RESULT 10

US-10-437-963-184040
; Sequence 184040, Application US/10437963
; Publication No. US20040123343A1

; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 184040
; LENGTH: 1446
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:


```
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46968
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-10-282-122A-46968

Query Match          26.5%; Score 58; DB 15; Length 296;
Best Local Similarity 36.2%; Pred. No. 15;
Matches 17; Conservative 5; Mismatches 21; Indels 4; Gaps 1;

QY 1 GLKLTVAASMDRVPKVTTPSSAISIARENHPEPL----GLNGIAE 43
   ||||| | | | | | | | | | | | | | | | | | | | |
Db 186 GLRLATTGAIKDSTINLQDVAVVALRFMEKHYPRLKRYNLNDIPE 232

RESULT 15
US-10-425-114-64239
; Sequence 64239, Application US/10425114
; Publication No. US2004003488A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 64239
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3606-012-B7_FLI.pep
US-10-425-114-64239

Query Match          26.5%; Score 58; DB 15; Length 391;
Best Local Similarity 37.8%; Pred. No. 22;
Matches 17; Conservative 5; Mismatches 11; Indels 12; Gaps 2;

QY 2 LKLTVAASMDRVPK-----VTPSSAISIARENHPEPR 35
   ||||| | | | | | | | | | | | | | | | | | | | |
Db 291 LKLTAAASCORPPQKMDQPGVAENISPSDACQDIS-QNVFPER 334

Search completed: November 11, 2004, 01:28:12
Job time : 49.392 secs
```

This Page Blank (uspto)

Result No.	Score	Query %		Length	DB	ID	Description
		Match	Ident				
1	213	97.3	147	2	T17303		hypothetical prote
2	62	28.3	471	2	S35635		DNA injection prot
3	61.5	28.1	1392	2	A43336		microtubule-vesicl
4	61.5	28.1	1437	2	S23695		restin - human
5	59	26.9	100	2	A83384		hypothetical prote
6	59	26.9	1146	2	F83853		S-methylcitrhydro
7	57	26.0	1911	2	T43048		calcium channel al
8	56	25.6	215	2	AD0695		ribonuclease T (im
9	55.5	25.3	1169	2	S31811		flocculation prote
10	55	25.1	378	2	H91032		probable DNA injec
11	55	25.1	378	2	A85877		probable prophage
12	55	25.1	413	1	S03631		homeotic protein S
13	55	25.1	903	2	JE0327		dynamidin-related pr
14	55	25.1	903	2	T00334		dynamidin-related pr
15	54.5	24.9	687	2	T09051		PepA protein - Pse
16	54.5	24.9	1693	1	MNWHE		genome polyprotein
17	54	24.7	780	1	S76487		probable copper-tr
18	54	24.7	1567	2	T00568		hypothetical prote
19	53.5	24.4	375	2	T51854		RING-H2 finger pro
20	53.5	24.4	656	2	D84359		thermosome subunit
21	52.5	24.0	2722	2	E88320		hypothetical prote
22	52.5	24.0	2738	2	F68320		protein F07A11.6 [
23	52	23.7	233	2	C71845		ferric transport A
24	52	23.7	382	2	F69880		conserved hypotnet
25	52	23.7	204	2	F72346		M-related protein
26	52	23.7	466	2	H71314		probable K+ transp
27	52	23.7	563	2	A69491		probable-formylmet
28	52	23.7	790	2	T27458		SMY2 protein - yea
29	52	23.7	898	2	T26577		hypothetical prote

RESULT 15
T09051
PePA protein - *Pseudomonas aeruginosa*
C.Species: *Pseudomonas aeruginosa*
C.Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C.Accession: T09051
R.Hauser, A.R.; Kang, P.J.; Engel, J.N.
Mol. Microbiol. 27, 807-818, 1998
A.Title: PePA, a secreted protein of *Pseudomonas aeruginosa* is necessary for cytotoxicity
A.Reference number: Z16536; MUID:98175457; PMID:9515706
A.Accession: T09051
A.Status: preliminary; translated from GB/EMBL/DDJJ
A.Molecule type: DNA
A.Residues: 1-687 >HAU>
A.Cross-references: UNIPROT:O34208; EMBL:AF027291; NID:G2967683; PIDN:AAC38269.1; PID:G2967683

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 54.9482 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-30

Perfect score: 219

Sequence: 1 GLKLTVAASMDRVPKVTSPS.....IAREHPEPRLGLNGIAETT 45

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	213	97.3	147	2 Q9UFM9	Q9ufm9 homo sapien
2	213	97.3	215	2 Q9NUT8	Q9nut8 homo sapien
3	213	97.3	676	2 Q7L9J1	Q7l9j1 homo sapien
4	213	97.3	763	2 Q96RR7	Q96rr7 homo sapien
5	211	96.3	727	2 Q80Y70	Q80y70 mus musculus
6	211	96.3	784	2 Q8BXC5	Q8bxc5 mus musculus
7	211	96.3	808	2 Q8C014	Q8c014 mus musculus
8	210	95.9	804	2 Q8DJR9	Q8djr9 xenopus tro
9	170.5	77.9	751	2 Q7ZUK6	Q7zuke brachydanio
10	96	43.8	596	2 Q9H7T7	Q9h7t7 homo sapien
11	96	43.8	742	2 Q8NAQ4	Q8naq4 homo sapien
12	96	43.8	763	2 Q96RR6	Q96rr6 homo sapien
13	96	43.8	763	2 Q92299	Q92299 mus musculus
14	96	43.8	813	2 Q8NE21	Q8ne21 homo sapien
15	96	43.8	813	2 Q8C9X6	Q8c9x6 mus musculus
16	96	43.8	836	2 Q9H2F5	Q9h2f5 homo sapien
17	67	30.6	478	2 Q76H14	Q76h14 salmonella
18	67	30.6	478	2 BAD15223	BAD15223 salmonella
19	66.5	30.4	684	2 Q7SEGI	Q7segi neurospora
20	64.5	29.5	1693	1 POLN_HEVPA	P33424 hepatitis e
21	64.5	29.5	1693	2 Q81344	Q81344 hepatitis e
22	64.5	29.5	1693	2 Q81862	Q81862 hepatitis e
23	64.5	29.5	1693	2 Q81876	Q81876 hepatitis e
24	64.5	29.5	1693	2 Q89444	Q89444 hepatitis e
25	64.5	29.5	1693	2 Q8UYD6	Q8uyd6 hepatitis e
26	62	28.3	471	1 VG20_BPP22	Q01076 bacterioph
27	62	28.3	471	2 Q77D87	Q77d87 bacterioph
28	62	28.3	471	2 AAM81396	AAM81396 bacteriop
29	61.5	28.1	653	2 Q8F529	Q8f529 homo sapien
30	61.5	28.1	653	2 AAR62543	AAR62543 homo sapi
31	61.5	28.1	1427	1 REST_HUMAN	P30622 homo sapien

ALIGNMENTS

RESULT 1

Q9UFM9 PRELIMINARY; PRT; 147 AA.
AC Q9UFM9; 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Hypothetical protein DKFZp566F2124 (Fragment).
GN Name=DKFZp566F2124;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Bloecker H., Boecher M., Brandt P., Mewes H.W., Gassenhuber J.,
RA Wiemann S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AL117558; CAB55993.1; -
DR FIR; T17303; T17303.
DR InterPro; IPR009607; E_Pc_C.
DR Pfam; PF06752; E_Pc_C_1.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 147 AA; 15246 MW; 4A6EF2AC8A83C323 CRC64;
Query Match 97.3%; Score 213; DB 2; Length 147;
Best Local Similarity 100.0%; Pred. No. 4.9e-20;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2

Q9NUT8 PRELIMINARY; PRT; 215 AA.
AC Q9NUT8; 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Hypothetical protein FLJ11148.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX PubMed=14702039;
RA Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H.,


```

OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RX MEDLINE=22389257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jorde L.B., Moore T., Max S.I., Wang J., Heide F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FVB/N; TISSUE=Colon;
RA Strausberg R.;
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC048785; AAH48785.1; -
DR MGD; MGI:1277221; D2Ertcd694e.
DR InterPro; IPR009607; E_Pc_C;
DR Pfam; PF06752; E_Pc_C; 1
FT NON_TER 1
SQ SEQUENCE 727 AA; 81796 MW; B9663DFBA912C3A6 CRC64;

Query Match 96.3%; Score 211; DB 2; Length 727;
Best Local Similarity 97.7%; Pred. No. 5e-19;
Matches 43; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 LKLTVAASMDRVPKVPFPSSAIISSIAIARENHEPERLGLNGIAETT 45
Db 678 LKLTVAASMDRVPKVPFPSSAIISSIAIARENHEPERLGLNGIAETT 721

RESULT 6
ID Q8BXCS PRELIMINARY; PRT; 784 AA.
AC Q8BXCS;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched
DE library, clone:CI3023C07 product:ENHANCER OF POLYCOMB 1 homolog.
GN Name=D2Ertcd694e;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RA "Functional annotation of a full-length mouse cDNA collection.";

```

```

RL Nature 409:695-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA The FANTOM Consortium;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RX MEDLINE=20309133; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:11757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Head;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK047947; BAC33197.1; -
DR MGD; MGI:1277221; D2Ertcd694e.
DR InterPro; IPR009607; E_Pc_C;
DR Pfam; PF06752; E_Pc_C; 1.
SQ SEQUENCE 784 AA; 88272 MW; 761FEEEC6F8BC47B CRC64;

Query Match 96.3%; Score 211; DB 2; Length 784;
Best Local Similarity 97.7%; Pred. No. 6e-19;
Matches 43; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 LKLTVAASMDRVPKVPFPSSAIISSIAIARENHEPERLGLNGIAETT 45
Db 735 LKLTVAASMDRVPKVPFPSSAIISSIAIARENHEPERLGLNGIAETT 778

RESULT 7
ID Q8C0I4 PRELIMINARY; PRT; 808 AA.
AC Q8C0I4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus adult male thymus cDNA, RIKEN full-length enriched
DE library, clone:583049L14 product:ENHANCER OF POLYCOMB 1 homolog.
GN Name=D2Ertcd694e;
OS Mus musculus (Mouse).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN NCBT_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.;
 RL "High-efficiency full-length cDNA cloning.";
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium;
 RL "Functional annotation of a full-length mouse cDNA collection.";
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RA The FANTOM Consortium;
 RA "The RIKEN Genome Exploration Research Group Phase I & II Team;
 "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=20493974; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu M., Hayashizaki Y.;
 RA Konno H., Okazaki Y., Muramatsu M., Sugahara Y., Shibata K., Itoh M.,
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes.";
 [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Azawa K., Nagao S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer.";
 [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Thymus;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Adachi J., Azawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
 RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
 RA Hayashida K., Hayatsu N., Hiramoto K., Hiracka T., Hirozane T.,
 RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa H., Kasukawa T.,
 RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
 RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
 RA Nishi K., Nomura K., Numazaki R., Ono M., Ohsato N., Okazaki Y.,
 RA Saito K., Saitoh H., Sakai K., Sakai K., Sakazume N., Sano H.,
 RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
 RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
 RA Tomaru A., Toyota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK031025; BAC27218.1;
 DR MGB: MGI:1277221; DZERC694e.
 DR InterPro: IPR009607; E_PC_C.
 DR Pfam: PF06752; E_PC_C; 1.
 SQ SEQUENCE 808 AA; 90987 MW; C11D11768C471DDB CRC64;

Query Match 96.3%; Score 211; DB 2; Length 808;
 Best Local Similarity 97.7%; Pred. No. 6.2e-19;
 Matches 43; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LKLTAAASMDRVPKVTSSAIISSIAIENHEPERLGLNGIAETT 45
 Db 759 LKLTAAASMDRVPKVTSSAIISSIAIENHEPERLGLNGIAETT 802
 [1]
 RESULT 8
 Q6DUR9 PRELIMINARY; PRT; 804 AA.
 ID Q6DUR9
 AC Q6DUR9
 DT 01-OCT-2004 (TrEMBLrel. 28, Created)
 DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Hypothetical protein.
 OS Xenopus tropicalis (Western clawed frog) (Silurana tropicalis).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 RN NCBT_TaxID=8364;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22389257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S., Loughran N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Maman A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywicki M.I., Skalska U., Small D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RA Klein S., Strausberg R.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC075105; AAH75105.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 804 AA; 90737 MW; FB320113258FAC2E CRC64;
 Query Match 95.9%; Score 210; DB 2; Length 804;
 Best Local Similarity 97.7%; Pred. No. 8.4e-19;
 Matches 43; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LKLTAAASMDRVPKVTSSAIISSIAIENHEPERLGLNGIAETT 45
 Db 755 LKLTAAASMDRVPKVTSSAIISSIAIENHEPERLGLNGIAETT 798
 [1]
 RESULT 9
 Q7ZUK6 PRELIMINARY; PRT; 751 AA.
 ID Q7ZUK6
 AC Q7ZUK6
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Similar to Enhancer of Polycomb.
 GN Name=zgc:55784;
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.

NCBI_TaxID=7955;
[1]
SEQUENCE FROM N.A.
STRAIN=AB; TISSUE=Whole body;
MEDLINE=22388257; PubMed=12477932;
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
Altshuler S.F., Zeeberg B., Buettow K.H., Schaefer C.B., Bhat N.K.,
Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
Stapleton M.J., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
Brownstein M., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Guaratne P.H.,
Richards S., Worley K.C., Hale S., Garcia A.M., Gay I.N., Hulyar S.W.,
Villalón D.K., Muzyk D.M., Sodergren E.J., Lu X., Gibbs R.A.,
Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
Rodríguez A.C., Grimwood J., Schmütz J., Myers R.M., Butterfield Y.S.,
Krzywinski M.I., Skalka U., Smalls D.E., Scherch A., Schein J.E.,
Jones S.-J., Maizumi M.A.;
"Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
SEQUENCE FROM N.A.
STRAIN=AB; TISSUE=Whole body;
Strausberg R.;
Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
EMBL; BC048890; AAH48890.1; -
InterPro: IPR009607; E PC C.
Pfam: PF06752; E PC C.1.
SEQUENCE 751 AA; 84613 MW; D1EEC81558F76D04 CRC64;

Query Match 77.9%; Score 170.5; DB 2; Length 751;
Best Local Similarity 79.5%; Pred. No. 1.4e-13;
Matches 35; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

QY 2 LKLATVAASMDRPVKTPSPSAISIARENEPERLINGIAETT 45
|||||
DB 703 LKLASVASSLD RPVKTPPTSAI-DIARENEPERALNGLSSET 745
|||||

RESULT 10
Q9HT7T PRELIMINARY; PRS; 596 AA.

ID Q9HT7T AC Q9HT7T DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TEMBLrel. 26, Last annotation update)
DE Hypothetical protein FLJ14267.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=Placenta;
PubMed=14702039;
Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
Wakanatsu A., Hayashi K., Satoh H., Nagai K., Kimura K., Makita H.,
Sekine M., Okayashiki M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
Yamamoto J., Saito K., Kawai T., Isoya Y., Nakamura Y., Negahari K.,
Murakami K., Saiyda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
Muro H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
Abe K., Kamihara K., Katsumi N., Sato K., Tanikawa M., Yamazaki M.,
Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K.,
Tanai H., Kimata M., Watanabe M., Hirooka S., Chiba Y., Ishida S.,
Ono Y., Takiguchi S., Watanabe S., Yoshida M., Horuta T., Kusano J.,
Kaneshiro K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
Togiwa S., Komai F., Hara R., Takeuchi K., Arita M., Imose N.,

[illegible]

RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,
 RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yanashita R.,
 RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegai T., Sugano S.,
 RT "Complete sequencing and characterization of 21,243 full-length human
 RT cDNAs.",
 RL Nat. Genet. 36:40-45(2004).
 DR EMBL; AK092304; BAC03857.1; -.
 DR InterPro: IPR009607; E_Pc_C_1.
 DR Pfam; PF06752; E_Pc_C_1.
 SQ SEQUENCE 742 AA; 82701 MW; 54707B36C4F8BF60 CRC64;

Query Match 43.8%; Score 96; DB 2; Length 742;
 Best Local Similarity 52.3%; Pred. No. 0.0011;
 Matches 23; Conservative 7; Mismatches 12; Indels 2; Gaps 1;

QY 2 LKLTVAASMDRVPKVTPTSSAIISSIAIARENHPEPRLGLNGIAETT 45
 Db 695 LKLA--AAANCQVSKVPSSSVDPRENHSEKFPALNNIADNT 736

RESULT 12

Q96RR6 PRELIMINARY; PRT; 763 AA.

AC Q96RR6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Enhancer of polycomb 1.
 GN Name=EPCL;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nunes D.N., Dias-Neto E., Brentani R.R., Camargo A.A.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF298905; AAK60501.1; -.
 DR InterPro: IPR009607; E_Pc_C_1.
 DR Pfam; PF06752; E_Pc_C_1.
 SQ SEQUENCE 763 AA; 85219 MW; 588D375F40151D03 CRC64;

Query Match 43.8%; Score 96; DB 2; Length 763;
 Best Local Similarity 52.3%; Pred. No. 0.0011;
 Matches 23; Conservative 7; Mismatches 12; Indels 2; Gaps 1;

QY 2 LKLTVAASMDRVPKVTPTSSAIISSIAIARENHPEPRLGLNGIAETT 45
 Db 716 LKLA--AAANCQVSKVPSSSVDPRENHSEKFPALNNIADNT 757

RESULT 13

Q92Z99 PRELIMINARY; PRT; 763 AA.

AC Q92Z99;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Enhancer of polycomb.
 GN Name=EPCL;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stankunas K., Berger J., Ruse C., Sinclair D.A., Randazzo F.,
 RA Brock H.W.;
 RT "The enhancer of polycomb gene of Drosophila encodes a chromatin
 RT protein conserved in yeast and mammals."
 RL Development 125:4055-4066(1998).
 DR EMBL; AF079765; AAC64272.1; -.

DR MGD; MGI:1278322; Epc1.
 DR InterPro: IPR009607; E_Pc_C_1.
 DR Pfam; PF06752; E_Pc_C_1.
 SQ SEQUENCE 763 AA; 84767 MW; D13649721FFB41E2 CRC64;

Query Match 43.8%; Score 96; DB 2; Length 763;
 Best Local Similarity 52.3%; Pred. No. 0.0011;
 Matches 23; Conservative 7; Mismatches 12; Indels 2; Gaps 1;

QY 2 LKLTVAASMDRVPKVTPTSSAIISSIAIARENHPEPRLGLNGIAETT 45
 Db 716 LKLA--AAANCQVSKVPSSSVDPRENHSEKFPALNNIADNT 757

RESULT 14

Q8NE21 PRELIMINARY; PRT; 813 AA.

AC Q8NE21;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE EPC1 protein.
 GN Name=EPCL;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Strausberg R.;
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC036529; AAK36529.1; -.
 DR InterPro: IPR009607; E_Pc_C_1.
 DR Pfam; PF06752; E_Pc_C_1.
 SQ SEQUENCE 813 AA; 90893 MW; 1A7DE9D14C291D11 CRC64;

Query Match 43.8%; Score 96; DB 2; Length 813;
 Best Local Similarity 52.3%; Pred. No. 0.0012;
 Matches 23; Conservative 7; Mismatches 12; Indels 2; Gaps 1;

QY 2 LKLTVAASMDRVPKVTPTSSAIISSIAIARENHPEPRLGLNGIAETT 45
 Db 766 LKLA--AAANCQVSKVPSSSVDPRENHSEKFPALNNIADNT 807

RESULT 15

Q8C9X6 PRELIMINARY; PRT; 813 AA.

AC Q8C9X6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A430080L23 product:enhancer of polycomb homolog 1,
DE (Drosophila), full insert sequence.
GN Name=Epc1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=92279253; PubMed=10349636;
RX Carninci P., Hayashizaki Y.;
RA "High-efficiency full-length cDNA cloning.";
RT Meth. Enzymol. 303:19-44(1999).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=21085660; PubMed=11217851;
RX RIKEN FANTOM Consortium;
RA "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690(2001).
RN [3]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Thymus;
RC The FANTOM Consortium;
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=20493374; PubMed=11042159;
RX Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RA "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Thymus;
RC MEDLINE=20530913; PubMed=11076861;
RX Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsumi T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RA "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=Thymus;
RC Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiroaka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akashira S., Takeda Y., Tanaka T.,
RA Tonaru A., Tota T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK040254; BAC30552.1;

DR MGD; MGI:1278322; Epc1.
DR InterPro: IPR009607; E_Pc_C.
DR Pfam: PF06752; E_Pc_C_1.
SQ SEQUENCE 813 AA; 90410 MW; E6F3CD987FC55905 CRC64;
Query Match 43.8%; Score 96; DB 2; Length 813;
Best Local Similarity 52.3%; Pred. No. 0.0012;
Matches 23; Conservative 7; Mismatches 12; Indels 1;
Gaps 1;
QY 2 LKLTATVAASMDRVPKVPSSAIISSIAIARENHHEPERLGLNGIAETT 45
DB 766 LKLA--AAACQVSKVPSSSVDPRENHSEKFPALNNIADNT 807
Search completed: November 10, 2004, 13:38:28
Job time : 57.9482 secs

This Page Blank (uspto)

EARLIER FILING DATE: 1998-05-18
NUMBER OF SEQ ID NOS: 39
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1261
TYPE: PRT
ORGANISM: Homo sapiens
US-09-566-076-2

Query Match 45.5%; Score 65; DB 4; Length 1261;
Best Local Similarity 44.0%; Pred. No. 0.17;
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MRDLPGHYETLKLVLGHLKTIADH 25
DB 807 LRQLPASNFSLHLFLVHLKRWVDH 831

RESULT 3
US-09-854-133-434
; Sequence 434, Application US/09854133
; Patent No. 6759508
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 434
; LENGTH: 1702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-854-133-434

Query Match 43.4%; Score 62; DB 4; Length 1702;
Best Local Similarity 44.0%; Pred. No. 0.69;
Matches 11; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 MRDLPGHYETLKLVLGHLKTIADH 25
DB 101 IQQLPPHYRTLEFLMRHLSLLADY 125

RESULT 4
US-09-248-796A-14141
; Sequence 14141, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14141
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14141

Query Match 39.2%; Score 56; DB 4; Length 422;

Best Local Similarity 42.3%; Pred. No. 1.2;
Matches 11; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
QY 1 MRDLPGHYETLKLVLGHLKTIADH 26
DB 385 LHSPLVVNFTKLVLKHLNKISEHK 410

RESULT 5
PCT-US93-03076-10
; Sequence 10, Application PC/TUS9303076
; GENERAL INFORMATION:
; APPLICANT: Whitehead Institute for Biomedical Research
; TITLE OF INVENTION: GAP-Associated Protein p190 and
; TITLE OF INVENTION: Transduction
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: 2 Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/03076
; FILING DATE: 19930331
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI92-03A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 165 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-03076-10

Query Match 37.8%; Score 54; DB 5; Length 165;
Best Local Similarity 45.5%; Pred. No. 0.84;
Matches 10; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 4 LPGHYETLKLVLGHLKTIADH 25
DB 99 LPPACETLRLVLAHLKRVTLH 120

RESULT 6
US-09-507-765-32
; Sequence 32, Application US/09507765
; Patent No. 6509155
; GENERAL INFORMATION:
; APPLICANT: Klingner, Tod M.
; APPLICANT: Stewart, Elizabeth A.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS
; FILE REFERENCE: PC-0010 US
; CURRENT APPLICATION NUMBER: US/09/507,765
; CURRENT FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 32
; LENGTH: 333


```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6509155 g6572185
US-09-507-765-32

Query Match      37.1%; Score 53; DB 4; Length 333;
Best Local Similarity 39.1%; Pred.No.2.7;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY      1 MRDLPGHYYETLKFVLVGHKLTIA 23
      :|||:|||:|||:|||:|||:
Db      207 LRSLPENYVVLRYLMGFLHAVS 229

RESULT 7
US-09-507-765-30
; Sequence 30, Application US/09507765
; Patent No. 6509155
; GENERAL INFORMATION:
; APPLICANT: Klinger, Tod M.
; APPLICANT: Stewart, Elizabeth A.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS
; FILE REFERENCE: PC-0010 US
; CURRENT APPLICATION NUMBER: US/09/507,765
; CURRENT FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 30
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6509155 3068538CD1
US-09-507-765-30

Query Match      37.1%; Score 53; DB 4; Length 433;
Best Local Similarity 39.1%; Pred.No.3.7;
Matches 9; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY      1 MRDLPGHYYETLKFVLVGHKLTIA 23
      :|||:|||:|||:|||:|||:
Db      307 LRSLPENYVVLRYLMGFLHAVS 329

RESULT 8
US-09-507-765-31
; Sequence 31, Application US/09507765
; Patent No. 6509155
; GENERAL INFORMATION:
; APPLICANT: Klinger, Tod M.
; APPLICANT: Stewart, Elizabeth A.
; APPLICANT: Yue, Henry
; APPLICANT: Baughn, Mariah R.
; TITLE OF INVENTION: GTPASE ACTIVATING PROTEINS
; FILE REFERENCE: PC-0010 US
; CURRENT APPLICATION NUMBER: US/09/507,765
; CURRENT FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 31
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6509155 404424.5.pseq
US-09-507-765-31

```

; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 51893
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-51893

Query Match 33.6%; Score 48; DB 4; Length 142;
Best Local Similarity 40.0%; Pred. No. 6.2;
Matches 8; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

QY 5 PGHYETLKFVGHKLTIADH 24
Db 55 PNHYPDLLRLRLGRNLKE 74

RESULT 12
US-07-841-997A-4
; Sequence 4, Application US/07841997A
; Patent No. 5422254
; GENERAL INFORMATION:
; APPLICANT: Lonsborough, John
; APPLICANT: Vuorio, Outi
; TITLE OF INVENTION: A method to increase the trehalose content
; TITLE OF INVENTION: of organisms by transforming them with the
; TITLE OF INVENTION: structural genes for the short and long chains
; TITLE OF INVENTION: yeast trehalose synthase.
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Alko Ltd.
; STREET: PO Box 350
; CITY: Helsinki
; STATE: Finland
; COUNTRY: Finland
; ZIP: SF-00101

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WP5.1 file exported as DOS text file
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/841,997A
; FILING DATE: 19920228
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/836,021
; FILING DATE: February 14, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary E. Gormley
; REGISTRATION NUMBER: 34409
; REFERENCE/DOCKET NUMBER: 920085A
; TELEPHONE: (202) 659-2930
; TELEFAX: (202) 887-0357
; TELEX: 440142
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 785 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: Linear
; MOLECULE TYPE: Polypeptide
; HYPOTHETICAL: Yes
; FRAGMENT TYPE: C-terminal
US-07-841-997A-4

Query Match 33.2%; Score 47.5; DB 1; Length 785;
Best Local Similarity 27.3%; Pred. No. 54;
Matches 12; Conservative 5; Mismatches 6; Indels 21; Gaps 2;

QY 4 LPHGYE-----TLKF-----LVGHLKTIADH 26

Db 604 LFGSYKINESMIKHTENAEDQDRVASVIGDAITHINTVFDHR 647

RESULT 13

US-08-290-301-4
; Sequence 4, Application US/08290301
; Patent No. 5792921
; GENERAL INFORMATION:
; APPLICANT: Lonsborough, John
; APPLICANT: Tunnela, Outi
; APPLICANT: Palva, Tupio
; APPLICANT: Holmstrom, Kjell-Ove
; APPLICANT: Welin, Bjorn
; APPLICANT: Mandel, Abul
; TITLE OF INVENTION: Increasing the trehalose content
; TITLE OF INVENTION: of organisms by transforming them with combinations of
; TITLE OF INVENTION: the structural genes for trehalose synthase.
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Alko Ltd.
; STREET: PO Box 350
; CITY: Helsinki
; STATE: Finland
; COUNTRY: Finland
; ZIP: SF-00101

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WP5.1 file exported as DOS text file
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/290,301
; FILING DATE: 15 August 1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FI 943133
; FILING DATE: 29 June 1994
; APPLICATION NUMBER: PCT/FI93/00049
; FILING DATE: 15 February 1993
; APPLICATION NUMBER: 07/841,997
; FILING DATE: 28 February 1992
; APPLICATION NUMBER: 07/836,021
; FILING DATE: 14 February 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Kubovcik, Ronald J.
; NAME: Lydon, James C.
; REGISTRATION NUMBER: 25,401
; REFERENCE/DOCKET NUMBER: 30,082
; TELEPHONE: (202) 467-6300
; TELEFAX: (202) 466-2006
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 785 amino acids
; TYPE: Amino acid
; TOPOLOGY: Linear
; MOLECULE TYPE: Polypeptide
; HYPOTHETICAL: Yes
; FRAGMENT TYPE: C-terminal
US-08-290-301-4

Query Match 33.2%; Score 47.5; DB 1; Length 785;
Best Local Similarity 27.3%; Pred. No. 54;
Matches 12; Conservative 5; Mismatches 6; Indels 21; Gaps 2;

QY 4 LPHGYE-----TLKF-----LVGHLKTIADH 26
Db 604 LFGSYKINESMIKHTENAEDQDRVASVIGDAITHINTVFDHR 647

RESULT 14

```
US-09-013-598-4
; Sequence 4, Application US/09013598
; Patent No. 6323001
; GENERAL INFORMATION:
; APPLICANT: Londerborough, John
; APPLICANT: Tunnela, Outi
; APPLICANT: Palva, Tupio
; APPLICANT: Holmstrom, Kjell-Ove
; APPLICANT: Welin, Bjorn
; APPLICANT: Mandel, Abul
; TITLE OF INVENTION: Increasing the trehalose content
; TITLE OF INVENTION: of organisms by transforming them with combinations of
; TITLE OF INVENTION: the structural genes for trehalose synthase.
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESS: PO Box 350
; CITY: Helsinki
; STATE: Finland
; COUNTRY: Finland
; ZIP: SF-00101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 720 KB
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WP5.1 file exported as DOS text file
; CURRENT APPLICATION NUMBER: US/09/013.598
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION NUMBER: 08/290,301
; FILING DATE:
; APPLICATION NUMBER: PCT/FI93/00049
; FILING DATE: 15 February 1993
; APPLICATION NUMBER: 07/841,997
; FILING DATE: 28 February 1992
; APPLICATION NUMBER: 07/836,021
; FILING DATE: 14 February 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Kubovcik, Ronald J.
; NAME: Lydon, James C.
; REGISTRATION NUMBER: 25,401
; REGISTRATION NUMBER: 30,082
; REFERENCE/DOCKET NUMBER: LAIN-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 467-6300
; TELEFAX: (202) 466-2006
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 785 amino acids
; TYPE: Amino acid
; TOPOLOGY: Linear
; MOLECULE TYPE: Polypeptide
; HYPOTHETICAL: Yes
; FRAGMENT TYPE: C-terminal
; US-09-013-598-4

Query Match 33.2%; Score 47.5; DB 3; Length 785;
Best Local Similarity 27.3%; Pred No. 54;
Matches 12; Conservative 5; Mismatches 6; Indels 21; Gaps 2;

Qy 4 LPHGYE---TLKF-----LVGHLKTIADHR 26
||| ||| : |||
Db 604 LPSYKINESMIKHTENAEDQDRVASVIGDAITHINTVDFHR 647

RESULT 15
US-08-290-301-82
; Sequence 82, Application US/08290301
; Patent No. 5792921
; GENERAL INFORMATION:
; APPLICANT: Londerborough, John
; APPLICANT: Tunnela, Outi
; APPLICANT: Palva, Tupio
; APPLICANT: Holmstrom, Kjell-Ove
; APPLICANT: Welin, Bjorn
; APPLICANT: Mandel, Abul
; TITLE OF INVENTION: Increasing the trehalose content
; TITLE OF INVENTION: of organisms by transforming them with combinations of
; TITLE OF INVENTION: the structural genes for trehalose synthase.
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESS: PO Box 350
; CITY: Helsinki
; STATE: Finland
; COUNTRY: Finland
; ZIP: SF-00101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 720 KB
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WP5.1 file exported as DOS text file
; CURRENT APPLICATION NUMBER: US/08/290,301
; FILING DATE: 15 February 1993
; APPLICATION NUMBER: 07/841,997
; FILING DATE: 28 February 1992
; APPLICATION NUMBER: 07/836,021
; FILING DATE: 14 February 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Kubovcik, Ronald J.
; NAME: Lydon, James C.
; REGISTRATION NUMBER: 25,401
; REGISTRATION NUMBER: 30,082
; REFERENCE/DOCKET NUMBER: LAIN-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 467-6300
; TELEFAX: (202) 466-2006
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 amino acids
; TYPE: Amino acid
; TOPOLOGY: Linear
; MOLECULE TYPE: Polypeptide
; HYPOTHETICAL: Yes
; FRAGMENT TYPE: yes
; US-08-290-301-82

Query Match 33.2%; Score 47.5; DB 1; Length 1098;
Best Local Similarity 27.3%; Pred. No. 79;
Matches 12; Conservative 5; Mismatches 6; Indels 21; Gaps 2;

Qy 4 LPHGYE---TLKF-----LVGHLKTIADHR 26
||| ||| : |||
Db 917 LPSYKINESMIKHTENAEDQDRVASVIGDAITHINTVDFHR 960

Search completed: November 10, 2004, 13:44:04
Job time : 9.981 secs
```

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 16:36:12 ; Search time 27.9309 Seconds
(without alignments)
328.807 Million cell updates/sec

Title: US-10-092-750-31

perfect score: 143
Sequence: 1 MRDLPGHYYETLKFVLVGHKLTIADHR 26

Scoring table: BLOSUM62
Gapen 10.0 : Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pcp.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pcp.*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp.*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	%			SUMMARIES			Description
	Score	Match	Length	DB	ID		
1	143	100.0	26	14	US-10-092-750-31	Sequence 31, App	
2	134	93.7	735	15	US-10-072-012-726	Sequence 726, App	
3	92	64.3	1173	15	US-10-072-012-725	Sequence 725, App	
4	92	64.3	1386	14	US-10-153-668-212	Sequence 212, App	
5	92	64.3	1386	15	US-10-072-012-724	Sequence 724, App	
6	92	64.3	1354	14	US-10-153-668-470	Sequence 470, App	
7	92	64.3	1355	15	US-10-072-012-278	Sequence 278, App	
8	85	59.4	816	14	US-10-104-047-2050	Sequence 2050, App	
9	65	45.5	1261	14	US-10-177-980-2	Sequence 2, App11	
10	65	45.5	1261	15	US-10-072-012-733	Sequence 733, App	
11	65	45.5	1261	16	US-10-648-593-192	Sequence 192, App	
12	62	43.4	1702	9	US-09-738-973-434	Sequence 434, App	
13	62	43.4	1702	9	US-09-854-133-434	Sequence 434, App	

14	62	43.-4	1702	14	US-10-144-649A-434	Sequence 434, App
15	58	40.6	2167	9	US-09-801-368-56	Sequence 56, Appl
16	58	40.6	2167	9	US-10-369-493-1773	Sequence 1773, Ap
17	55	38.5	2935	14	US-10-205-194-95	Sequence 95, Appl
18	55	38.5	903	16	US-10-408-765A-827	Sequence 827, Appl
19	55	38.5	1021	15	US-10-188-248-24	Sequence 24, Appl
20	55	38.5	1126	15	US-10-108-260A-3665	Sequence 3665, Ap
21	53	37.1	170	9	US-09-802-127-7	Sequence 7, Appl
22	53	37.1	170	13	US-10-080-960-32	Sequence 32, Appl
23	53	37.1	333	14	US-10-284-753-32	Sequence 32, Appl
24	53	37.1	390	14	US-10-097-340-31	Sequence 31, Appl
25	53	37.1	433	14	US-10-284-753-30	Sequence 30, Appl
26	53	37.1	433	14	US-10-284-753-31	Sequence 31, Appl
27	53	37.1	464	14	US-10-097-340-14	Sequence 14, Appl
28	53	37.1	643	14	US-10-097-340-16	Sequence 16, Appl
29	53	37.1	718	14	US-10-097-340-19	Sequence 19, Appl
30	53	37.1	751	14	US-10-097-340-18	Sequence 18, Appl
31	52	36.4	468	15	US-10-113-944-258	Sequence 258, App
32	52	36.4	1101	9	US-09-823-950A-2	Sequence 2, Appl
33	51	35.7	103	13	US-10-080-960-34	Sequence 34, Appl
34	51	35.7	152	14	US-10-138-585-5	Sequence 5, Appl
35	50	35.0	523	15	US-10-072-012-732	Sequence 732, App
36	49	34.3	156	15	US-10-424-599-152761	Sequence 152761,
37	49	34.3	175	15	US-10-072-012-878	Sequence 878, App
38	49	34.3	175	15	US-10-072-012-879	Sequence 879, App
39	49	34.3	184	15	US-10-424-599-150592	Sequence 150592,
40	49	34.3	256	15	US-10-424-599-152765	Sequence 152765,
41	49	34.3	262	15	US-10-424-599-152762	Sequence 152762,
42	49	34.3	322	15	US-10-072-012-384	Sequence 284, App
43	49	34.3	403	15	US-10-424-599-150595	Sequence 150595,
44	49	34.3	981	15	US-10-424-599-150591	Sequence 150591,
45	48	33.6	445	15	US-10-424-599-178810	Sequence 178810,

ALIGNMENTS

```

RESULT 1
US-10-092-750-31
; Sequence 31, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-XL
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 26
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-31

```

```

; APPLICANT: Tchernev, Velizar
; APPLICANT: Spvtek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsbrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Rieger, Daniel K.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 726
; LENGTH: 735
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-726

Query Match 93.7%; Score 134; DB 15; Length 735;
Best Local Similarity 96.0%; Pred. No. 2.7e-11;
Matches 24; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MRDLPGHVYETLKFVLGHLKTIADH 25
:|||||:|||||:|||||:|||||:|||||:
DB 612 IRDLPGHVYETLKFVLGHLKTIADH 636
:|||||:|||||:|||||:|||||:|||||:

RESULT 3
US-10-072-012-725
; Sequence 725, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spvtek, Kimberly
; APPLICANT: Zerhusen, Bryan

```

```

; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsbrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Rieger, Daniel K.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 725
; LENGTH: 1173
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-725

Query Match 64.3%; Score 92; DB 15; Length 1173;
Best Local Similarity 64.0%; Pred. No. 9.3e-05;
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MRDLPGHVYETLKFVLGHLKTIADH 25
:|||||:|||||:|||||:|||||:|||||:
DB 478 IHDLPFHVYETLKFVLSAHLKTIADH 502
:|||||:|||||:|||||:|||||:|||||:

RESULT 4
US-10-153-668-212
; Sequence 212, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P

```

```

; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; PRIOR FILING DATE: 2001-10-10
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 212
; LENGTH: 1286
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-212

```

```

Query Match      64.3%; Score 92; DB 14; Length 1286;
Best Local Similarity 64.0%; Pred. No. 0.0001;
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      1 MRDLPGHYYETLKFLVGHKLTIADH 25
Db      591 IHDLPEHHYETLKFLSAHLKTVAE 615

```

```

RESULT 5
US-10-072-012-724
; Sequence 724, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spvtek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsbrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31

```

```

; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 724
; LENGTH: 1286
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-724

```

```

Query Match      64.3%; Score 92; DB 15; Length 1286;
Best Local Similarity 64.0%; Pred. No. 0.0001;
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      1 MRDLPGHYYETLKFLVGHKLTIADH 25
Db      591 IHDLPEHHYETLKFLSAHLKTVAE 615

```

```

RESULT 6
US-10-153-668-470
; Sequence 470, Application US/10153668
; Publication No. US20030092616A1
; GENERAL INFORMATION:
; APPLICANT: HONDA, Goichi
; APPLICANT: MATSUDA, Akio
; APPLICANT: MURAMATSU, Shuji
; APPLICANT: ISHIZAWA, Kenya
; TITLE OF INVENTION: STAT6 Activating Gene
; FILE REFERENCE: 1254-0207P
; CURRENT APPLICATION NUMBER: US/10/153,668
; CURRENT FILING DATE: 2002-05-24
; PRIOR APPLICATION NUMBER: US 60/293,172
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: US 60/316,031
; PRIOR FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 60/328,403
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: JP 2001-157043
; PRIOR FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: JP 2001-260681
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: JP 2001-313175
; NUMBER OF SEQ ID NOS: 488
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 470
; LENGTH: 1354
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-153-668-470

```

```

Query Match      64.3%; Score 92; DB 14; Length 1354;
Best Local Similarity 64.0%; Pred. No. 0.0001;
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```

```

QY      1 MRDLPGHYYETLKFLVGHKLTIADH 25
Db      659 IHDLPEHHYETLKFLSAHLKTVAE 683

```

```

RESULT 7
US-10-072-012-278
; Sequence 278, Application US/10072012

```

Publication No. US20040033493A1
GENERAL INFORMATION:
APPLICANT: Tchernev, Velizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zerhusen, Bryan
APPLICANT: Patturajan, Meera
APPLICANT: Shimkets, Richard
APPLICANT: Li, Li
APPLICANT: Gangolli, Esha
APPLICANT: Padigar, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Colman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A.
APPLICANT: Furtak, Katarzyna
APPLICANT: Grosse, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 278
LENGTH: 1355
TYPE: PRT
ORGANISM: Homo sapiens
US-10-072-012-278

Query Match 64.3%; Score 92; DB 15; Length 1355;
Best Local Similarity 64.0%; Pred. No. 0.00011;
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 MRDLPCHYVETLKFLVGHKLTIADH 25
DB 659 IHDLPEHHYETLKFLSAHLKTV 683

RESULT 8
US-10-104-047-2050
Sequence 2050, Application US/10104047
Publication No. US20030236392A1
GENERAL INFORMATION:
APPLICANT: HELIX RESEARCH INSTITUTE

TITLE OF INVENTION: No. US20030236392A1e1 full length cDNA
FILE REFERENCE: H1-A0105
CURRENT APPLICATION NUMBER: US/10/104,047
CURRENT FILING DATE: 2002-03-25
PRIOR APPLICATION NUMBER:
PRIOR FILING DATE:
NUMBER OF SEQ ID NOS: 4096
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2050
LENGTH: 816
TYPE: PRT
ORGANISM: Homo sapiens
US-10-104-047-2050

Query Match 59.4%; Score 85; DB 14; Length 816;
Best Local Similarity 68.2%; Pred. No. 0.0007;
Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 MRDLPCHYVETLKFLVGHKLTI 22
DB 795 IHDLPEHHYETLKFLSAHLKTV 816

RESULT 9

US-10-177-980-2
Sequence 2, Application US/10177980
Publication No. US20030166232A1
GENERAL INFORMATION:
APPLICANT: Saras, Jan
APPLICANT: Franz, Petra
APPLICANT: Aspenstrm, Pontus
APPLICANT: Hellman, Ulf
APPLICANT: Genez, Leonel Jorge
APPLICANT: Heldin, Carl-Henrik
TITLE OF INVENTION: PARG, A GTPASE ACTIVATING PROTEIN WHICH INTERACTS WITH PTPL1
FILE REFERENCE: L0461/7030
CURRENT APPLICATION NUMBER: US/10/177,980
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US/09/080,855
PRIOR FILING DATE: 1998-05-18
PRIOR APPLICATION NUMBER: 08/805,583
PRIOR FILING DATE: 1997-02-25
NUMBER OF SEQ ID NOS: 39
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 1261
TYPE: PRT
ORGANISM: Homo sapiens
US-10-177-980-2

Query Match 45.5%; Score 65; DB 14; Length 1261;
Best Local Similarity 44.0%; Pred. No. 1.1;
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 1 MRDLPCHYVETLKFLVGHKLTIADH 25
DB 807 LRQLPASNFNSLHFLVHLKRVVDH 831

RESULT 10

US-10-072-012-733
Sequence 733, Application US/10072012
Publication No. US20040033493A1
GENERAL INFORMATION:
APPLICANT: Tchernev, Velizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zerhusen, Bryan
APPLICANT: Patturajan, Meera
APPLICANT: Shimkets, Richard
APPLICANT: Li, Li
APPLICANT: Gangolli, Esha
APPLICANT: Padigar, Muralidhara
APPLICANT: Anderson, David W.


```
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 733
; LENGTH: 1261
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-733

Query Match      45.5%; Score 65; DB 15; Length 1261;
Best Local Similarity 44.0%; Pred. No. 1.1;
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy      1 MRDLPGHYYETLKLVLGHKLTIADH 25
Db      807 LRQLPASNFNSLHFLVLHKLKRVVDH 831

RESULT 11
US-10-648-593-192
; Sequence 192, Application US/10648593
; Publication No. US20040106132A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
; TITLE OF INVENTION: INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE PATHWAYS IN BREAST CELLS
; FILE REFERENCE: D0273 NP
; CURRENT APPLICATION NUMBER: US/10/648,593
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: 60/406,385
; PRIOR FILING DATE: 2002-08-27
; NUMBER OF SEQ ID NOS: 557
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 192
```

```
; LENGTH: 1261
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-648-593-192

Query Match      45.5%; Score 65; DB 16; Length 1261;
Best Local Similarity 44.0%; Pred. No. 1.1;
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

Qy      1 MRDLPGHYYETLKLVLGHKLTIADH 25
Db      807 LRQLPASNFNSLHFLVLHKLKRVVDH 831

RESULT 12
US-09-738-973-434
; Sequence 434, Application US/09738973
; Patent No. US20020110563A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Henderson, Robert A.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Fling, Steven P.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Algate, Paul A.
; APPLICANT: Secrist, Heather
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Benson, Darin R.
; APPLICANT: Eliot, Mark
; APPLICANT: Mannion, Jane
; APPLICANT: Kalos, Michael D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C9
; CURRENT APPLICATION NUMBER: US/09/738,973
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 587
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 434
; LENGTH: 1702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-973-434

Query Match      43.4%; Score 62; DB 9; Length 1702;
Best Local Similarity 44.0%; Pred. No. 4.5;
Matches 11; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Qy      1 MRDLPGHYYETLKLVLGHKLTIADH 25
Db      101 IQQLPPPHVYRTLEFLMRHLSLLADY 125

RESULT 13
US-09-854-133-434
; Sequence 434, Application US/09854133
; Publication No. US20020183499A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Henderson, Robert A.
; APPLICANT: Benson, Darin R.
; APPLICANT: Secrist, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: THE THERAPY AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.475C10
; CURRENT APPLICATION NUMBER: US/09/854,133
; CURRENT FILING DATE: 2001-05-11
; NUMBER OF SEQ ID NOS: 735
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 434
; LENGTH: 1702
; TYPE: PRT
```

```
; ORGANISM: Homo sapiens
US-09-854-133-434

Query Match      43.4%; Score 62; DB 9; Length 1702;
Best Local Similarity 44.0%; Pred. No. 4.5;
Matches 11; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY      1 MRDLPGHYETLKFVLVGHKLTIADH 25
       ::|||::|||::|||::|||::|||:
DB      101 IQQLPPPHRYTLEFLMRHLSLLADY 125
       ::|||::|||::|||::|||::|||:

RESULT 14
US-10-144-649A-434
; Sequence 434, Application US/10144649A
; Publication No. US20030118599A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Algate, Paul A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; FILE REFERENCE: 210121.475C11
; CURRENT APPLICATION NUMBER: US/10/144.649A
; CURRENT FILING DATE: 2002-08-21
; NUMBER OF SEQ ID NOS: 749
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 434
; LENGTH: 1702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-144-649A-434

Query Match      43.4%; Score 62; DB 14; Length 1702;
Best Local Similarity 44.0%; Pred. No. 4.5;
Matches 11; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY      1 MRDLPGHYETLKFVLVGHKLTIADH 25
       ::|||::|||::|||::|||::|||:
DB      101 IQQLPPPHRYTLEFLMRHLSLLADY 125
       ::|||::|||::|||::|||::|||:

RESULT 15
US-09-801-368-56
; Sequence 56, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Milne, Todd
; APPLICANT: Maxon, Mary
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801.368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56

; LENGTH: 2167
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-56

Query Match      40.6%; Score 58; DB 9; Length 2167;
Best Local Similarity 44.0%; Pred. No. 24;
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY      1 MRDLPGHYETLKFVLVGHKLTIADH 25
       ::|||::|||::|||::|||::|||:
DB      2090 LQKLTCTYQTLKRIVFHLNKHQH 2114
       ::|||::|||::|||::|||::|||:

Search completed: November 11, 2004, 01:28:13
Job time : 28.9809 secs
```

GenCore version S.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 14:52:32 ; Search time 5.12456 Seconds
(without alignments)
281.634 Million cell updates/sec

Title: US-10-092-750-228

Perfect score: 86

Sequence: 1 TLRHGGLQFNTRFGV 15.

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match: 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	57.0	540	2 B40901	equine arteritis v
2	44	51.2	277	2 T40174	conserved hypothet
3	44	51.2	363	2 S57763	branched-chain ami
4	43	50.0	98	2 A64143	hypothetical prote
5	43	50.0	128	2 T26721	hypothetical prote
6	43	50.0	307	2 A42902	transcription regu
7	43	50.0	307	2 F83198	transcription regu
8	42	48.8	650	2 T23175	hypothetical prote
9	41.5	48.3	412	2 D95941	conserved hypothet
10	41.5	48.3	412	2 AF3448	hypothetical membr
11	41	47.7	147	2 JC2302	ougl protein - Bac
12	41	47.7	265	2 H95140	transporter, PNT f
13	41	47.7	265	2 H98008	formate-nitrate tr
14	41	47.7	277	2 B72226	hypothetical prote
15	41	47.7	285	2 AF0146	conserved hypothet
16	41	47.7	346	2 JC1376	alcohol dehydrogen
17	41	47.7	346	2 H70860	alcohol dehydrogen
18	40.5	47.1	49	2 A90016	hypothetical prote
19	40	46.5	275	2 S77009	hypothetical prote
20	40	46.5	283	2 AE2346	hypothetical prote
21	40	46.5	467	2 AC2015	hypothetical prote
22	40	46.5	509	2 T45529	agaA protein [limp
23	40	46.5	1463	2 A36861	orf 1b protein - L
24	39	45.3	415	2 PL0165	hemocyanin chain b
25	39	45.3	492	2 AB1013	probable membrane
26	39	45.3	525	2 T30345	irp5 protein - Yec
27	39	45.3	525	2 AG0232	yersiniabactin sid
28	39	45.3	525	2 T17443	salicyl-AMP ligase
29	39	45.3	657	1 BHLOA	hemocyanin chain a

RESULT 1

B40901

equine arteritis virus lb protein homolog - lactic dehydrogenase virus (isolate LDV-P) (f

C:Species: lactic dehydrogenase virus

C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 09-Jul-2004

C:Accession: B40901

R:Kuo, L.; Harty, J.T.; Erickson, L.; Palmer, G.A.; Plagemann, P.G.W.

J. Virol. 65, 5118-5123, 1991

A:Title: A nested set of eight RNAs is formed in macrophages infected with lactate dehydri

A:Reference number: A40901; MUID:91333066; PMID:1870216

A:Accession: B40901

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-540 <KIO>

A:Cross-references: UNIPROT:O83024; GB:S50064; MID:G233624; PIDN:AA19478.1; PID:G233625

C:Superfamily: equine arteritis virus RNA-directed RNA polymerase

Query Match 57.0%; Score 49; DB 2; Length 540;

Best Local Similarity 75.0%; Pred. No. 2.7;

Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 3 RHWGLQFNTRFG 14

DB 50 RHRGVLFNTRFG 61

RESULT 2

T40174

conserved hypothetical protein SPBC30B4.08 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C:Accession: T40174

R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Hilbert, H.; Duesterhoeft, A.

submitted to the EMBL Data Library, August 1998

A:Reference number: Z21909

A:Accession: T40174

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-277 <LYN>

A:Cross-references: UNIPROT:O74367; EMBL:AL031262; PIDN:CAA20321.1; GSPDB:GN00067; SPDB:1

A:Experimental source: strain 972h-; cosmid c30B4

C:Genetics:

A:Gene: SPDB:SPBC30B4.08

A:Map position: 2

A:Introns: 43/2; 178/3

Query Match 51.2%; Score 44; DB 2; Length 277;

Best Local Similarity 87.5%; Pred. No. 8.9;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LRHWGLQF 9

DB 2 LRHWGLQF 9

hemocyanin chain b
probable MYB trans
hypothetical prote
ionB-dependent rec
probable polyketid
endoglucanase homo
alpha-hemolysin BH
hypothetical prote
14K beta-galactosi
hydrolase [impor
sucrose-phosphatas
hypothetical prote
probable regulator
cellulase - Fibrob
glycyl-tRNA synthe
arginine-tRNA liga

ALIGNMENTS

Db 216 LEHWGLOF 223

RESULT 3

S57763

branched-chain amino acid aminotransferase homolog ywaA - Bacillus subtilis

C:Species: Bacillus subtilis

C>Date: 19-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004

C:Accession: S57763; S39655; A70050

R:Glaser, P.; Lubochinsky, B.; Danchin, A.

submitted to the EMBL Data Library, July 1995

A:Description: Bacillus subtilis cel operon.

A:Reference number: S57758

A:Accession: S57763

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-363 <GLA>

A:Cross-references: UNIPROT:P39576; EMBL:Z49992; NID:g895746; PIDN:CAA90289.1; PID:g8957

R:Glaser, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionescu, A.; Rapoport, G.; Danchin, A.

Mol. Microbiol. 10, 371-384, 1993

A:Title: Bacillus subtilis genome project: cloning and sequencing of the 97 kb region fr

A:Reference number: S39655; MUID:95020537; PMID:7934828

A:Accession: S39655

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-58 <GL2>

A:Cross-references: EMBL:X73124; NID:g413923; PIDN:CAA51556.1; PID:g413924

R:Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter

C: Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, V.; Capuano, V.; Carter, N.M.; Chd

A: Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holzapfel, S.; Hosono, S.; Hullo, M.F.

Koeter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapius, A.; Lardinois

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle

Rieger, M.; Rivolta, C.; Roche, B.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron

teuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstake, P.; Tognoni, K.; Tosato, V.; Uchiyama,

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: A70050

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-363 <KUN>

A:Cross-references: GB:299123; GB:AL009126; NID:g2636240; PIDN:CAB15881.1; PID:g2636390

A:Experimental source: strain 168

C:Genetics:

A:Gene: ywaA

C:Superfamily: branched-chain-amino-acid transaminase BA11

Query Match 51.2%; Score 44; DB 2; Length 363;

Best Local Similarity 63.6%; Pred. No. 12;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 LRHWGLOFNTFR 12

Db 269 LKHGWLQVSR 279

RESULT 4

A64143

hypothetical protein HI0112 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 09-Jul-2004

C:Accession: A64143

R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, C

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: A64143

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-98 <TIG>

A:Cross-references: UNIPROT:P71339; GB:U32696; GB:L42023; NID:g1573057; PIDN:AAC21791.1;

A>Note: Best homolog was a hypothetical protein (insertion element IS1223) from Lactobaci

Query Match 50.0%; Score 43; DB 2; Length 98;

Best Local Similarity 80.0%; Pred. No. 4.5;

Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TLRHWGLOFN 10

Db 55 TLRHWNOFN 64

RESULT 5

T36721

hypothetical protein Y39A1A.20 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T36721

R:Wall, M.

submitted to the EMBL Data Library, September 1998

A:Reference number: Z20257

A:Accession: T36721

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-128 <WIL>

A:Cross-references: UNIPROT:Q9XX30; EMBL:AL031633; PIDN:CAA21010.1; GSPDB:GN00021; CESP:1

A:Experimental source: clone Y39A1A

C:Genetics:

A:Gene: CESP:Y39A1A.20

A:Map position: 3

A:introns: 33/2; 47/1; 77/3; 107/1

Query Match 50.0%; Score 43; DB 2; Length 128;

Best Local Similarity 72.7%; Pred. No. 5.9;

Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 RHWGLOFNTRF 13

Db 67 RQWMLQFYTRF 77

RESULT 6

A42902

transcription regulator mmsR - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa

C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004

C:Accession: A42902; S27601

R:Steele, M.I.; Lorenz, D.; Hatter, K.; Park, A.; Sokatch, J.R.

J. Biol. Chem. 267, 13585-13592, 1992

A:Title: Characterization of the mmsAB operon of Pseudomonas aeruginosa PAO encoding met

A:Reference number: A42902; MUID:92317087; PMID:1339433

A:Accession: A42902

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-307 <STE>

A:Cross-references: UNIPROT:P28809; GB:M84911; NID:g151360; PIDN:AAA25890.1; PID:g151361

A:Experimental source: PAO, ATCC 15692

A>Note: Sequence extracted from NCBI backbone (NCBIN:107704, NCBI:P:107706)

C:Genetics:

A:Gene: mmsR

C:Superfamily: hypothetical protein b2382

C:Keywords: DNA binding; transcription regulation

Query Match 50.0%; Score 43; DB 2; Length 307;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRHWGLQ 8
|||||
Db 144 LRHWGLQ 150

RESULT 7
F83198
transcription regulator MmrR PA3571 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: F83198
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: F83198
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-307 <STO>
A:Cross-references: UNIPROT:P28809; GB:AE004778; GB:AE004091; NID:99949722; PIDN:AAG0695
A:Experimental source: strain PA01
C:Genetics:
A:Gene: mmsR; PA3571
C:Superfamily: hypothetical protein b2382

Query Match 50.0%; Score 43; DB 2; Length 307;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LRHWGLQ 8
|||||
Db 144 LRHWGLQ 150

RESULT 8
T23175
hypothetical protein K01C8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T23175
R:Sims, M.
Submitted to the EMBL Data Library, April 1995
A:Reference number: Z19702
A:Accession: T23175
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-650 <WIL>
A:Cross-references: UNIPROT:Q21087; EMBL:Z49066; PIDN:CAA88862.1; GSPDB:GN00020; CESP:K01C8
A:Experimental source: clone K01C8
C:Genetics:
A:Gene: CESP:K01C8.3
A:Map position: 2
A:Introns: 31/1; 74/1; 181/3; 392/2; 416/1; 489/3; 551/1
F:78-546/Domain: animal histidine decarboxylase homology <HDC>

Query Match 48.8%; Score 42; DB 2; Length 650;
Best Local Similarity 54.5%; Pred. No. 45;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 3 RHWGLQFTRF 13
||||| : |||
Db 422 RHWGLPLSRF 432

RESULT 9
D95941
conserved hypothetical membrane protein, paralogue of Y20848 Smb21292 [imported] - Sinorhizobium meliloti
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
C:Accession: D95941

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmaster, J.; Chain, P.; Vorholter, F.J.; Hernanc Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmaid from the N2-fixing endosymbiont of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: D95941
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-412 <KUR>
A:Cross-references: UNIPROT:Q92VB2; GB:AL591985; PIDN:CAC49186.1; PID:GL5140681; GSPDB:G092VB2
A:Experimental source: strain 1021, megaplasmaid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubier, P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.
Science 293, 658-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, M.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.C.
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Smb21292
A:Genome: plasmid

Query Match 48.3%; Score 41.5; DB 2; Length 412;
Best Local Similarity 66.7%; Pred. No. 34;
Matches 8; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

QY 1 LRHWGLQFTR 12
|||||
Db 266 TIRHW---FNTR 274

RESULT 10
AF3448
hypothetical membrane spanning protein BMEI1572 [imported] - Brucella melitensis (strain 19/54)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AF3448
R:DeiVecchio, V.G.; Kaputal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, C.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, C.
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3448
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-417 <KUR>
A:Cross-references: UNIPROT:Q8YFF0; GB:AE008917; PIDN:AAL52753.1; PID:gl7983585; GSPDB:G08917
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI1572
A:Map position: 1

Query Match 48.3%; Score 41.5; DB 2; Length 417;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 8; Conservative 1; Mismatches 0; Indels 3; Gaps 1;

QY 1 LRHWGLQFTR 12
|||||
Db 266 TIRHW---FNTR 274

RESULT 11
JC2302
ougl protein - Bacillus licheniformis
C:Species: Bacillus licheniformis
C>Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
C:Accession: JC2302
R:Yoshida, K.; Seki, S.; Fujita, Y.
DNA Res. 1, 157-162, 1994
A:Title: Nucleotide sequence and features of the Bacillus licheniformis gnt operon.
A:Reference number: PC2201; MUID:96051988; PMID:8535972
A:Accession: JC2302

```

A:Molecule type: DNA
A:Residues: 1-147 <YOS>
A:Cross-references: UNIPROT:Q45292; DBJ:D31631; NID:9563948; PIDN:BA06500.1; PID:95639
C:Genetics:
A:Gene: oug
C:Superfamily: Bacillus licheniformis oug protein

  Query Match      47.7%; Score 41; DB 2; Length 147;
  Best Local Similarity 66.7%; Pred. No. 14;
  Matches 8; Conservative 0; Mismatches 2; Indels 2; Gaps 1;

Qy 3 LRHWGLQFNTRFG 14
|||||:|
Db 119 RHWGYSF--RFG 128

RESULT 12
H95140
transporter, FMT family, probable SPI215 [imported] - Streptococcus pneumoniae (strain T
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C:Accession: H95140
R:Tectelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.B.; Lewis, M.R.; Radune, D.; Holtzapfle,
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95140
A>Status: preliminary
A:Molecule type: DNA
A:Cross-references: 1-265 <KUR>
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI215

  Query Match      47.7%; Score 41; DB 2; Length 265;
  Best Local Similarity 59.3%; Pred. No. 26;
  Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 LRHWGLQFNTRF 13
|||||:|
Db 231 LRHWGVTFIGNF 242

RESULT 13
H98008
formate-nitrate transporter [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C:Accession: H98008
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: H98008
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-265 <KUR>
A:Cross-references: UNIPROT:Q97QJ1; UNIPROT:Q8DPM4; GB:AE007317; PIDN:AAK99900.1; PID:91
C:Genetics:
A:Gene: nlrC

  Query Match      47.7%; Score 41; DB 2; Length 265;
  Best Local Similarity 58.3%; Pred. No. 26;
  Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 LRHWGLQFNTRF 13
|||||:|
Db 231 LRHWGVTFIGNF 242

RESULT 14
B72226
hypothetical protein - Thermotoga maritima (strain MSBs)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 09-Jul-2004
C:Accession: B72226
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwin, M.L.; Dodson, R.J.; Haft, D.H.; Hickey,
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: B72226
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-277 <ARN>
A:Cross-references: UNIPROT:Q9X204; GB:AE001808; GB:AE000512; NID:54982233; PIDN:AAD3674;
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1676

  Query Match      47.7%; Score 41; DB 2; Length 277;
  Best Local Similarity 50.0%; Pred. No. 28;
  Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 5 WGLQFNTRFG 14
|||||:|
Db 161 WGLEYSKFG 170

RESULT 15
AF0146
conserved hypothetical protein YP01192 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: AF0146
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; E
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, E
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AF0146
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-285 <KUR>
A:Cross-references: UNIPROT:Q8ZGT5; GB:AL590842; PIDN:CAC90033.1; PID:gl5979255; GSPDB:GN
C:Genetics:
A:Gene: YP01192

  Query Match      47.7%; Score 41; DB 2; Length 285;
  Best Local Similarity 46.2%; Pred. No. 28;
  Matches 6; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 2 LRHWGLQFNTRFG 14
|||||:|
Db 87 VHRWGTPTFNAYG 99

Search completed: November 10, 2004, 15:55:10
Job time : 5.12456 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 14:50:40 ; Search time 25.1423 Seconds
(without alignments)
343.270 Million cell updates/sec

Title: US-10-092-750-228
Perfect score: 86
Sequence: 1 TLRHGLQFNTRFGV 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 02:.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	57.0	79	Q6JHZ7	Q6jhz7 homo sapien
2	49	57.0	540	Q83024	Q83024 lactate deh
3	49	57.0	882	Q7SKF3	Q7skf3 human immun
4	49	57.0	1410	Q83018	Q83018 lactate deh
5	47	54.7	319	Q8Y066	Q8y066 raietonia s
6	47	54.7	326	1 HEMZ_PASMU	P57874 pasteurella
7	46	53.5	352	Q9UIFO	Q9uifo leishmania
8	45	52.3	189	Q8A8V1	Q8a8v1 bacteroides
9	45	52.3	645	Q7URC4	Q7urc4 rhodopirell
10	44.5	51.7	2325	Q9N3X8	Q9n3x8 caenorhabdi
11	44	51.2	231	Q13658	Q13658 schizosacch
12	44	51.2	277	Q74367	Q74367 schizosacch
13	44	51.2	362	1 ILVE_BACSU	P39576 bacillus su
14	44	51.2	402	Q6FJN1	Q6fjn1 candida gla
15	44	51.2	779	Q7RWK9	Q7rkw9 plasmodium
16	44	51.2	1038	Q7PPW7	Q7ppw7 anopheles g
17	43	50.0	98	Y112_HAEIN	P71339 haemophilus
18	43	50.0	196	Q89TR3	Q89tr3 bradyrhizob
19	43	50.0	247	Q9ANL4	Q9anl4 bradyrhizob
20	43	50.0	285	1 ALKE_BABBO	P40690 babesia bov
21	43	50.0	307	1 MMSR_PSEAE	P28809 pseudomonas
22	43	50.0	308	Q82PA3	Q82pa3 streptomyc
23	43	50.0	308	Q9L149	Q9l149 streptomyc
24	43	50.0	319	Q7MYE6	Q7mye6 photorhabdu
25	43	50.0	323	Q93H27	Q93h27 streptomyc
26	43	50.0	426	1 CINA_GLOVI	Q7nnu4 glucosebacter
27	43	50.0	740	Q82MNS	Q82mns nitrosomona
28	43	50.0	1057	Q7V0Y9	Q7v0y9 prochloroco
29	42	48.8	144	Q7MYB5	Q7myb5 photorhabdu
30	42	48.8	380	Q8S0V9	Q8s0v9 oryza sativ
31	42	48.8	650	Q21087	Q21087 caenorhabdi

RESULT 1
Q6JHZ7
ID Q6JHZ7 PRELIMINARY; PRT; 79 AA.
AC Q6JHZ7;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE HCV-NS5A1P5 binding protein 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Zhang J., Cheng J., Wang L., Shao Q., Lu Y., Chen T., Hong Y.;
RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY459291; AAR23236.1;
SQ SEQUENCE 79 AA; 8682 MW; 3023D182934D2618 CRC64;

Query Match 57.0%; Score 49; DB 2; Length 79;
Best Local Similarity 80.0%; Pred. No. 1.5;
Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 2
Q83024
ID Q83024 PRELIMINARY; PRT; 540 AA.
AC Q83024;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Equine arteritis virus 1b protein homolog protein (Fragment).
OS Lactate dehydrogenase-elevating virus (LDV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Arteriviridae; Arterivirus.
OX NCBI_TaxID=11048;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=9133066; PubMed=1870216;
RA Xuo L.L., Hartley J.T., Erickson L., Palmer G.A., Plagemann P.G.;
RT "A nested set of eight RNAs is formed in macrophages infected with
RT lactate dehydrogenase-elevating virus.";
RL J. Virol. 65:5118-5123(1991).
DR EMBL; S50064; AAB19478.1;
DR EMBL; S50064; AAB19478.1;
DR EMBL; S50064; AAB19478.1;
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR InterPro; IPR010459; Corona_RPol.
DR InterPro; IPR007095; RNA_pol_DS.
DR InterPro; IPR007094; RNA_pol_PSVir.

Q95ZS2 caenorhabdi
Q6kaf3 oryza sativ
Bad19111 oryza sat
Q9u5d7 plautia sta
Q92vb2 rhizobium m
Q8g2g9 brucella su
Q9yff0 brucella me
Q45292 bacillus li
Q8dpm4 streptococc
Q97qj1 streptococc
Q9x204 thermotoga
Q8zgt5 versinia pe
Aas61198 versinia
P31975 mycobacteri

32 42 48.8 705 2 Q95ZS2
33 42 48.8 750 2 Q6KAF3
34 42 48.8 750 2 BAD19111
35 42 48.8 1856 2 Q9U5D7
36 41.5 48.3 412 2 Q92VB2
37 41.5 48.3 414 2 Q8G2G9
38 41.5 48.3 417 2 Q9YFF0
39 41 47.7 147 1 YOUNG_BACLI
40 41 47.7 265 2 Q8DPM4
41 41 47.7 265 2 Q97QJ1
42 41 47.7 277 2 Q9X204
43 41 47.7 285 2 Q8ZGT5
44 41 47.7 285 2 AAS61198
45 41 47.7 346 1 ADHC_MYCTU

ALIGNMENTS

DR InterPro; IPR001607; Znf_UBP.
 DR Pfam; PF06332; Corona_RPOL; 1.
 DR Pfam; PF02148; zf-UBP; 1.
 FT NON_TER 540 540
 SQ SEQUENCE 540 AA; 60507 MW; A51F2A424FFB974E CRC64;
 Query Match 57.0%; Score 49; DB 2; Length 540;
 Best Local Similarity 75.0%; Pred. No. 11;
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 3 RHWGLQFNTRFG 14
 DB 50 RHGVLFNTRFG 61
 RESULT 3
 Q7SKF3 PRELIMINARY; PRT; 882 AA.
 AC Q7SKF3;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Envelope glycoprotein.
 GN Name=env;
 OS Human immunodeficiency virus 2.
 OC Viruses; Retroviridae; Retroviridae; Lentivirus.
 ON NCBI_TaxID=11709;
 RN SEQUENCE FROM N.A.
 RP Bhanja P., Mandal D., Chakrabarti S.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY309063; AAQ19802.1; .
 DR GO; GO:0016021; C: integral to membrane; IEA.
 DR GO; GO:0019028; C: viral capsid; IEA.
 DR GO; GO:0019031; C: viral envelope; IEA.
 DR GO; GO:0005198; F: structural molecule activity; IEA.
 DR InterPro; IPR000328; Env_GP41.
 DR InterPro; IPR000777; GP120.
 DR Pfam; PF00516; GP120; 1.
 DR Pfam; PF00517; GP41; 1.
 KW AIDS; Coat protein; Envelope protein; Glycoprotein; Polyprotein; Transmembrane.
 SQ SEQUENCE 882 AA; 101055 MW; 3274398DE6451D82 CRC64;
 Query Match 57.0%; Score 49; DB 2; Length 882;
 Best Local Similarity 61.5%; Pred. No. 19;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 2 LRHWGLQFNTRFG 14
 DB 594 LNHWGLSLKKFG 606
 RESULT 4
 ID Q83018 PRELIMINARY; PRT; 1410 AA.
 AC Q83018;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 25, Last annotation update)
 DE Polyprotein 1b (Fragment).
 OS Lactate dehydrogenase-elevating virus (LDV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
 OC Arteriviridae; Arterivirus.
 ON NCBI_TaxID=11048;
 RN SEQUENCE FROM N.A.
 RP STRAIN=Plagemann;
 RC MEDLINE=95297165; PubMed=7778295;
 RA Palmer G.A., Kuo L., Chen Z., Faaberg K.S., Plagemann P.G.;
 RT "Sequence of the genome of lactate dehydrogenase-elevating virus: heterogeneity between strains P and C.";
 RL Virology 209:637-642(1995).

RN SEQUENCE FROM N.A.
 RP STRAIN=Plagemann;
 RC MEDLINE=96010224; PubMed=7571421;
 RX Faaberg K.S., Plagemann P.G.;
 RA "The envelope proteins of lactate dehydrogenase-elevating virus and their membrane topography.";
 RT Virology 212:512-525(1995).
 RL EMBL; U15146; AAA85664.1; .
 DR PIR; J01998; J01998.
 DR GO; GO:0003723; F: RNA binding; IEA.
 DR GO; GO:0003724; F: RNA helicase activity; IEA.
 DR GO; GO:0003968; F: RNA-directed RNA polymerase activity; IEA.
 DR GO; GO:0019079; F: viral genome replication; IEA.
 DR InterPro; IPR010459; Corona_RPOL.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR InterPro; IPR000506; Viral_helicase.
 DR Pfam; PF06332; Corona_RPOL; 1.
 DR Pfam; PF01443; Viral_helicase; 1.
 DR Pfam; PF02148; zf-UBP; 1.
 KW Polyprotein.
 FT NON_TER 1
 SQ SEQUENCE 1410 AA; 155812 MW; DB6BEA525BA6B82B CRC64;
 Query Match 57.0%; Score 49; DB 2; Length 1410;
 Best Local Similarity 75.0%; Pred. No. 31;
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 3 RHWGLQFNTRFG 14
 DB 150 RHGVLFNTRFG 161
 RESULT 5
 Q8Y066 PRELIMINARY; PRT; 319 AA.
 ID Q8Y066;
 AC Q8Y066;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE PROBABLE LIPOPROTEIN TRANSMEMBRANE.
 GN Name=RSC04543; OrderedLocustNames=RSC1178;
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 ON NCBI_TaxID=305;
 RN SEQUENCE FROM N.A.
 RP STRAIN=GMI1000;
 RC MEDLINE=21681879; PubMed=11823852;
 RX Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S., Ariat M., Sillault A., Brotter P., Camus J.C., Cattolico L., Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N., Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T., Siguier P., Thebault P., Whalen M., Wincker P., Levy M., Weissbach J., Boucher C.A.;
 RA "Genome sequence of the plant pathogen Ralstonia solanacearum.";
 RT Nature 415:497-502(2002).
 RL EMBL; AL646063; CAD14880.1; .
 DR GO; GO:0003824; F: catalytic activity; IEA.
 DR GO; GO:0045735; F: nutrient reservoir activity; IEA.
 DR InterPro; IPR002641; Patatin.
 DR Pfam; PF01734; Patatin; 1.
 KW Complete proteome; Lipoprotein; Transmembrane.
 SQ SEQUENCE 319 AA; 32967 MW; 6924E74D4F53342 CRC64;
 Query Match 54.7%; Score 47; DB 2; Length 319;
 Best Local Similarity 57.1%; Pred. No. 14;
 Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
 QY 1 LRHWGLQFNTRFG 14


```

Db      117 TIADWALPFQTRFG 130
|: | | | | | | |
RESULT 6
HEMZ_PASMU STANDARD; PRT; 326 AA.
ID      HEMZ_PASMU STANDARD; PRT; 326 AA.
AC      P57874;
AD      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      05-JUL-2004 (Rel. 44, Last annotation update)
DE      Ferrochelatase (EC 4.99.1.1) (Protoheme ferro-lyase) (Heme
DE      synthetase)
DE      Name=NhmH; OrderedLocusNames=PM0789;
OS      Pasteurella multocida.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC      Pasteurellaceae; Pasteurella.
OX      NCBI_TaxID=747;
ON      [1]
RN      SEQUENCE FROM N.A.
RP      STRAIN=PM70;
RC      MEDLINE=211145866; PubMed=11248100; DOI=10.1073/pnas.051634598;
RX      May B.-J., Zhang Q., Li L.B., Paustean M.L., Whittam T.S., Kapur V.;
RT      "Complete genomic sequence of Pasteurella multocida PM70."
RL      Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
CC      -1- FUNCTION; Catalyzes the ferrous insertion into protoporphyrin IX.
CC      -1- CATALYTIC ACTIVITY; Protoporphyrin + Fe(2+) = protoheme + 2 H(+).
CC      -1- PATHWAY; Protoheme biosynthesis; last step.
CC      -1- SUBCELLULAR LOCATION; Cytoplasmic (by similarity).
CC      -1- SIMILARITY; Belongs to the ferrochelatase family.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sb-sib.ch).
CC      -----
CC      EMBL; AE006117; AA02873.1; -.
DR      HAMAP; MF_00323; -. 1
DR      InterPro; IPR001015; Ferrochelatase.
DR      Pfam; PF00762; Ferrochelatase; 1.
DR      ProDom; PD002792; Ferrochelatase; 1.
DR      TIGRFAMs; TIGR00109; hemH; 1.
DR      PROSITE; PS00534; FERROCHELATASE; 1.
KW      Complete proteome; Heme biosynthesis; Iron; Lyase;
KW      Porphyrin biosynthesis.
FT      METAL 195 195 Iron (By similarity).
FT      METAL 276 276 Iron (By similarity).
FT      FT
SQ      SEQUENCE 326 AA; 37421 MW; BAAAEERAC77DE3D5 CRC64;
Query Match 54.7%; Score 47; DB 1; Length 326;
Best Local Similarity 57.1%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy      1 TLRHWGLQFNTRFG 14
| | | | |
Db      226 TENQWGLSFQSRFG 239
| | | | |

RESULT 7
Q9U1FO PRELIMINARY; PRT; 352 AA.
ID      Q9U1FO PRELIMINARY; PRT; 352 AA.
AC      Q9U1FO;
AD      01-MAY-2000 (TrEMBLrel. 13, Created)
DT      01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT      01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE      NADP-dependent alcohol hydrogenase.
DE      Name=Li063.09;
OS      Leishmania major.
OS      Leishmania major.
OC      Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX      NCBI TaxID=55664;
ON      NCBI TaxID=55664;

```

```

Db      68 LRHWGIQPEVAV 79
|||||:|
|:
RESULT 9
QURC4
ID Q7URC4 PRELIMINARY; PRT; 645 AA.
AC Q7URC4;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Probable chaperone protein Dnak.
GN Name=dnaK; OrderedLocusNames=RB5754;
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RA MEDLINE=22735913; PubMed=12835416;
RX Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
CC -!- SIMILARITY: Belongs to the heat shock protein 70 family.
DR EMBL; BX294142; CAD74416.1; -.
DR GO; GO:0005524; F-ATP binding; IEA.
DR InterPro; IP3001023; Hsp70.
DR Pfam; PF00012; HSP70; 1.
DR PROSITE; PS00297; HSP70_1; 1.
DR PROSITE; PS00329; HSP70_2; 1.
KW ATP-binding; Complete proteome.
SQ SEQUENCE 645 AA; 70160 MW; 2F23DBF17291F2C1 CRC64;

Query Match 52.3%; Score 45; DB 2; Length 645;
Best Local Similarity 70.0%; Pred. No. 65;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 3 RHWGLQFNTR 12
|:|:|:|
Db 631 RHWKLEFNAR 640
|:|:|:|

RESULT 10
Q9N3X8 PRELIMINARY; PRT; 2325 AA.
AC Q9N3X8;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Hypothetical protein Y42H9B.2.
GN ORFNames=Y42H9B.2;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA Wilson R.;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Tin-Wollam A., Wohldmann P.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]

Db      93 LRHWAASSPDVTKHEVDGTTNFTIDGLQPSRYGV 1030
|||||:|
|:|:|:|
Query Match 51.7%; Score 44.5; DB 2; Length 2325;
Best Local Similarity 28.9%; Pred. No. 3e+02;
Matches 11; Conservative 3; Mismatches 1; Indels 23; Gaps 1;

Qy 1 TLRW-----GLQFNTRFGV 15
|:|:|:|
Db 933 TIRHWAASSPDVTKHEVDGTTNFTIDGLQPSRYGV 1030
|:|:|:|

RESULT 11
O13668 PRELIMINARY; PRT; 231 AA.
ID O13668;
AC O13668;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Pi077 protein.
GN Name=pi077;
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972 h-;

```

```

RX MEDLINE=20089027; PubMed=10620777;
RA Machida M., Yamazaki S., Kunihito S., Tanaka T., Kushida N., Jinno K.,
RA Haikawa Y., Yamazaki J., Yamamoto S., Sekine M., Oguchi A., Nagai Y.,
RA Sakai M., Aoki K., Ogura K., Kudoh Y., Kikuchi H., Zhang M.Q.,
RA Vanagida M.;
RT "A 38 kb segment containing the cdc2 gene from the left arm of fission
RT yeast chromosome II: sequence analysis and characterization of the
RT genomic DNA and cDNAs encoded on the segment.";
RL Yeast 16:71-80(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972 h-;
RA Director-General of Biotechnology Center, Kushida N., Machida M.;
RL Submitted (Mar-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB004539; BAA21459.1; -
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0004527; F:exonuclease activity; IEA.
DR InterPro; IPR006055; Exonuclease.
DR Pfam; PF00929; Exonuc X-T; 1.
DR SMART; SM00479; EXOIII; 1.
SQ SEQUENCE 231 AA; 27071 MW; F804B2ED24F4CF54 CRC64;

Query Match 51.2%; Score 44; DB 2; Length 231;
Best Local Similarity 87.5%; Pred. No. 33;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LRHWGLQF 9
DB 170 LEHWGLQF 177

RESULT 12
O74367 PRELIMINARY; PRT; 277 AA.
ID O74367
AC O74367
DT 01-NOV-1998 (TrEMBLrel. 05, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE KapD family exonuclease; no apparent S. cerevisiae ortholog.
GN Name=SPBC30B4.08;
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972h-;
RX MEDLINE=21848401; PubMed=11899360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S., Warren T., Whitehead S.,
RA Woodward J., Voickert G., Art R., Robben J., Grymonprez B.,
RA Welljens I., Vanstreels E., Rieger M., Schafer M., Muller-Auerl S.,
RA Gabel C., Fuchs M., Dusterhoft A., Fritz C., Holzer E., Moestl D.,
RA Hilbert H., Borzym K., Langer I., Beck A., Lehach H., Reinhardt R.,
RA Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA Dominguez A., Revuelta J.L., Moreno S., Sanchez M., del Rey P., Benito J.,
RA Caput L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;

```

```

RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
DR EMBL; AL031262; CAA20321.1; -
DR PIR; T40174; T40174.
DR GeneDB SPombe; SPBC30B4.08; -
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0004527; F:exonuclease activity; IEA.
DR InterPro; IPR006055; Exonuclease.
DR InterPro; IPR003034; SAP.
DR Pfam; PF00929; Exonuc X-T; 1.
DR SMART; SM00479; EXOIII; 1.
DR PROSITE; PS50800; SAP; 1.
DR Exonuclease 277 AA; 32282 MW; 79CDAFFDC954C89 CRC64;
SQ SEQUENCE 277 AA; 32282 MW; 79CDAFFDC954C89 CRC64;

Query Match 51.2%; Score 44; DB 2; Length 277;
Best Local Similarity 87.5%; Pred. No. 40;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LRHWGLQF 9
DB 216 LEHWGLQF 223

RESULT 13
ILVE_BACSU STANDARD; PRT; 362 AA.
ID ILVE_BACSU
AC P39576;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Putative branched-chain-amino-acid aminotransferase (EC 2.6.1.42)
DE (BCAT) (Vegetative protein 85) (VEG85).
GN Name=ywaA; Synonyms=ipa-0r; OrderedLocNames=BSU38550;
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Glaser P., Lubochinsky B., Danchin A.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377; DOI=10.1038/36786;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bartoro M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.-D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M.,
RA Jones L.-M., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M.,
RA Klein C., Kobayashi Y., Koetter P., Koningstein G., Krogh S.,
RA Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J.,
RA Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maue C.,
RA Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S.,
RA Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudaga B.,
RA Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S.,
RA Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G.,
RA Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B.,
RA Rose M., Sadale Y., Sato T., Scanlan E., Schleich S., Schroeter R.,
RA Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Serror P.,
RA Shin B.S., Solido B., Sorokin A., Tacconi E., Takagi T., Takahashi H.,
RA Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P.,
RA Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F.,
RA Vassarotti A., Viari A., Wambut R., Wedler E., Wedler H.,

```

RA Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K.,
RA Yoshimoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumstein E.,
RA Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 1-57 FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7334828;
RA Glaeser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W., Hullo M.F.,
RA Ionescu M., Lubochinsky B., Marcelino L., Moszer I., Presecan E.,
RA Santana M., Schneider E., Schweizer J., Vertes A., Rapoport G.,
RA Danchin A.;
RT *Bacillus subtilis* genome project: cloning and sequencing of the 97 kb
RT region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
RN [4]
RP SEQUENCE OF 1-9.
RC STRAIN=168 / IS58;
RX MEDLINE=97443988; PubMed=9298659;
RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U., Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
RT regulation map for *Bacillus subtilis*.";
RL Electrophoresis 18:1451-1463(1997).
CC -!- FUNCTION: Acts on leucine, isoleucine and valine (By similarity).
CC -!- CATALYTIC ACTIVITY: L-leucine + 2-oxoglutarate = 4-methyl-2-
CC oxopentanoate + L-glutamate.
CC -!- COFACTOR: Pyridoxal phosphate.
CC -!- PATHWAY: Valine and isoleucine biosynthesis.
CC -!- SIMILARITY: Belongs to the Class-IV pyridoxal-phosphate-dependent
CC aminotransferase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z49992; CAA90289.1; -;
DR EMBL; Z99123; CAB15881.1; -;
DR EMBL; X73124; CAA51556.1; -;
DR PIR; S57763; S57763.
DR HSSP; O15382; 1KT8.
DR Subtilist; BG10546; ywaA.
DR InterPro; IPR001544; Aminotrans IV.
DR InterPro; IPR005786; B amino transII.
DR Pfam; PF01063; Aminotran 4; 1.
DR ProDom; PD001961; Aminotran 4; 1.
DR TIGRFAMs; TIGR01123; ilvE II; 1.
DR PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
DR AminoTransferase; Branched-chain amino acid biosynthesis;
KW Complete proteome; Direct protein sequencing; Pyridoxal phosphate;
KW Transferase.
FT INIT MET 0
FT BINDING 196 196 Pyridoxal phosphate (By similarity).
SQ SEQUENCE 362 AA; 40191 MW; FPBA681A5FDC7057 CRC64;
Query Match 51.2%; Score 44; DB 1; Length 362;
Best Local Similarity 53.6%; Pred. No. 52;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Oy 2 LRHWGLQFNTR 12
Db 268 LKHGLQVSR 278
RESULT 14
Q6FJN1
ID Q6FJN1 PRELIMINARY; PRT; 402 AA.
AC Q6FJN1;

DT 05-JUL-2004 (TRENBLrel. 27, Created)
DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)
DE Similar to sp|P11745|Saccharomyces cerevisiae YMR235C RNA1 GTPase
DE activating protein.
DE ORFNames=CAGLQMC4961G;
GN Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS138;
RG GENOLEVURES;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talia E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleykasten C.,
RA Boistrame A., Boyer J., Cattolico L., Confanieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicoud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellens S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Sweeney D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissbach J.,
RA Wincker P., Souciet J.L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
DR EMBL; CB380959; CAG52539.1; -;
DR InterPro; IPR001611; LRR.
DR InterPro; IPR003590; LRR_RNinh_sub.
DR Pfam; PF00560; LRR; 4.
DR PRINTS; PR00019; LEURICHRPT.
DR SMART; SM00368; LRR_R1; 5.
SQ SEQUENCE 402 AA; 45455 MW; 34C30B376A5AA95F CRC64;
Query Match 51.2%; Score 44; DB 2; Length 402;
Best Local Similarity 46.7%; Pred. No. 58;
Matches 7; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
Oy 1 TLRWGLQFNTRFGV 15
Db 221 TLHYGLKYNKLEI 235
RESULT 15
Q7RKW9
ID Q7RKW9 PRELIMINARY; PRT; 779 AA.
AC Q7RKW9;
DT 01-MAR-2004 (TRENBLrel. 26, Created)
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
DE 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Aldo-keto reductase.
GN Name=PY02780;
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=73239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=17XNL;
RX PubMed=12368865;
RA Carlton J.M., Angioli S.V., Suh B.B., Koolij T.W., Pertea M.,
RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,
RA Peterson J.D., Pop M., Kosack D.S., Shumway M.P., Bidwell S.L.,
RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblum T.V.,
RA Cho J.K., Quackenbush J., Sedegah M., Shoabai A., Cummings L.M.,
RA Florens L., Yates F.R. III, Raine J.D., Sinden R.S., Harris M.A.,
RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,
RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,
RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,
RA Carucci D.J.;

```

RT "Genome sequence and comparative analysis of the model rodent malaria
  parasite Plasmodium yoelii yoelii."
RL Nature 419:512-519(2002).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data
DR EMBL; AABL01000772; EAA22267.1; -.
DR InterPro; IPR001395; Aldo/ket_red.
DR Pfam; PF00248; Aldo_ket_red; 1.
DR PRINTS; PR00069; ALDKETRDPTASE.
DR ProDom; PD000288; Aldo/ket_red; 2.
SQ SEQUENCE 779 AA; 93034 MW; 44083D42DE648A00 CRC64;

Query Match      51.2%; Score 44; DB 2; Length 779;
Best Local Similarity 57.1%; Pred.No. 1.2e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      2 LRHWGLQFNTRFGV 15
Db      595 IREWGLSNETPFGV 608

```

Search completed: November 10, 2004, 15:53:45
 Job time : 27.2193 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:41:17 ; Search time 8.04167 Seconds
(without alignments)
222.664 Million cell updates/sec

Title: US-10-092-750-22

Perfect score: 140
Sequence: 1 ARGDFAAQQLWLRALGRPLPTSH 27

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:**

- 1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:**
- 2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:**
- 3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:**
- 4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:**
- 5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:**
- 6: /cgn2_6/prodata/1/1aa/backfiles1.pep:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	140	100.0	1147	1	US-08-131-365B-38
2	140	100.0	1147	2	US-08-668-123-38
3	75	53.6	629	3	US-09-360-237-60
4	75	53.6	1049	3	US-09-522-666-6
5	75	53.6	1141	1	US-08-131-365B-54
6	75	53.6	1141	2	US-08-668-123-54
7	48.5	34.6	503	4	US-09-252-991A-18345
8	48	34.3	478	4	US-09-252-991A-25034
9	48	34.3	849	4	US-09-252-991A-17953
10	48	34.3	988	4	US-09-382-552-233
11	48	34.3	2080	4	US-09-382-552-2
12	47	33.6	501	4	US-09-252-991A-31885
13	47	33.6	639	1	US-08-466-390-2
14	47	33.6	639	1	US-08-470-950-2
15	47	33.6	639	1	US-08-467-781-2
16	47	33.6	639	1	US-08-195-487-2
17	47	33.6	639	2	US-08-483-924-2
18	47	33.6	639	5	PCT-US93-06160-2
19	46.5	33.2	125	4	US-09-270-767-37479
20	46.5	33.2	125	4	US-09-270-767-52696
21	46	32.9	294	2	US-09-489-039A-12085
22	46	32.9	594	2	US-08-785-310A-6
23	46	32.9	940	4	US-09-252-991A-24639
24	45.5	32.5	476	4	US-09-252-991A-32962
25	45	32.1	366	4	US-09-489-039A-7237
26	45	32.1	429	4	US-09-252-991A-32753
27	45	32.1	435	4	US-09-328-352-5030

28	45	32.1	605	4	US-09-252-991A-25512	Sequence 25512, A
29	45	32.1	760	4	US-09-252-991A-18711	Sequence 18711, A
30	44.5	31.8	98	4	US-09-252-991A-23171	Sequence 23171, A
31	44.5	31.8	148	4	US-09-252-991A-16863	Sequence 16863, A
32	44.5	31.8	384	4	US-09-252-991A-16983	Sequence 16983, A
33	44	31.4	192	4	US-09-485-632B-2	Sequence 2, Appli
34	44	31.4	217	4	US-09-252-991A-23222	Sequence 23222, A
35	44	31.4	221	4	US-09-229-583A-4	Sequence 4, Appli
36	44	31.4	221	4	US-10-187-904-4	Sequence 4, Appli
37	44	31.4	251	4	US-09-632-287A-4	Sequence 4, Appli
38	44	31.4	251	4	US-09-131-237C-20	Sequence 20, Appli
39	44	31.4	312	3	US-09-252-991A-18082	Sequence 18082, A
40	44	31.4	7257	3	US-08-335-409-5	Sequence 5, Appli
41	44	31.4	7257	3	US-09-568-102-5	Sequence 5, Appli
42	44	31.4	7257	3	US-09-567-969-5	Sequence 5, Appli
43	44	31.4	7257	3	US-09-568-480-5	Sequence 5, Appli
44	44	31.4	7257	3	US-09-568-486-5	Sequence 5, Appli
45	44	31.4	7257	3	US-09-568-472-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-131-365B-38
; Sequence 38, Application US/08131365B
; Patent No. 5527690
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/131,365B
; FILING DATE: 01-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1147 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-131-365B-38

Query Match 100.0%; Score 140; DB 1; Length 1147;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ARGDFAAQQLWLRALGRPLPTSH 27
|||||

Db 594 ARGDFAQAQQLWLALRALGRPLPTSH 620

US-08-668-123-38

RESULT 2

US-08-668-123-38

Sequence 38, Application US/08668123

Patent No. 5891631

GENERAL INFORMATION:

APPLICANT: Brown, Michael S.

APPLICANT: Briggs, Michael R.

APPLICANT: Wang, Xiaodong

APPLICANT: Goldstein, Joseph L.

TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING

TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING

TITLE OF INVENTION: PROTEINS

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: U.S.A.

ZIP: 77210

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/668,123

FILING DATE: 14-JUN-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/131,365

FILING DATE: 01-OCT-1993

ATTORNEY/AGENT INFORMATION:

NAME: Parker, David L.

REGISTRATION NUMBER: 32,165

REFERENCE/DOCKET NUMBER: UTSD:372/PAR

TELECOMMUNICATION INFORMATION:

TELEPHONE: (512) 418-3000

TELEFAX: (512) 474-7577

INFORMATION FOR SEQ ID NO: 38:

SEQUENCE CHARACTERISTICS:

LENGTH: 1147 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-668-123-38

Query Match 100.0%; Score 140; DB 2; Length 1147;

Best Local Similarity 100.0%; Pred. No. 4.2e-12;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ARGDFAQAQQLWLALRALGRPLPTSH 27

Db 594 ARGDFAQAQQLWLALRALGRPLPTSH 620

RESULT 3

US-09-360-237-60

Sequence 60, Application US/09360237

Patent No. 632962

GENERAL INFORMATION:

APPLICANT: BROWN, MICHAEL S.

APPLICANT: CHENG, DONG

APPLICANT: ESPENSHADE, PETER J.

APPLICANT: GOLDSTEIN, JOSEPH L.

APPLICANT: RAWSON, ROBERT B.

APPLICANT: SAKAI, JURO

TITLE OF INVENTION: STEROL-REGULATED SITE-1 PROTEASE AND ASSAYS OF

TITLE OF INVENTION: MODULATORS THEREOF

FILE REFERENCE: UTXD:567

US-09-360-237-60

Query Match 53.6%; Score 75; DB 3; Length 629;

Best Local Similarity 65.4%; Pred. No. 0.0075;

Matches 17; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ARGDFAQAQQLWLALRALGRPLPTS 26

Db 74 ARGDFAAAAANLQCLAVLGRALEPTS 99

RESULT 4

US-09-522-666-6

Sequence 6, Application US/09522666

Patent No. 6333167

GENERAL INFORMATION:

APPLICANT: Shuey, David

APPLICANT: Quinet, Elaine

TITLE OF INVENTION: Methods and Reagents for Identifying Inhibitors of

TITLE OF INVENTION: Proteolysis of Membrane-Associated Proteins

FILE REFERENCE: 6-00

CURRENT APPLICATION NUMBER: US/09/522,666

CURRENT FILING DATE: 2000-03-10

NUMBER OF SEQ ID NOS: 32

SOFTWARE: Patent In Ver. 2.0

SEQ ID NO 6

LENGTH: 1049

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Amino acid

OTHER INFORMATION: sequence encoded from KpnI to XhoI in ISRP-laci

OTHER INFORMATION: Fusion Protein

US-09-522-666-6

Query Match 53.6%; Score 75; DB 3; Length 1049;

Best Local Similarity 65.4%; Pred. No. 0.013;

Matches 17; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 1 ARGDFAQAQQLWLALRALGRPLPTS 26

Db 492 ARGDFAAAAANLQCLAVLGRALEPTS 517

RESULT 5

US-08-131-365B-54

Sequence 54, Application US/08131365B

Patent No. 5527690

GENERAL INFORMATION:

APPLICANT: Brown, Michael S.

APPLICANT: Briggs, Michael R.

APPLICANT: Wang, Xiaodong

APPLICANT: Goldstein, Joseph L.

TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING

TITLE OF INVENTION: TO STEROL REGULATORY ELEMENT BINDING

TITLE OF INVENTION: PROTEINS

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: Arnold, White & Durkee

STREET: P.O. Box 4433

CITY: Houston

STATE: Texas

COUNTRY: U.S.A.


```
/
/
/ ZIP: 77210
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/131,365B
/ FILING DATE: 01-OCT-1993
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Parker, David L.
/ REGISTRATION NUMBER: 32,165
/ REFERENCE/DOCKET NUMBER: UTSD:372/PAR
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 54:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1141
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-131-365B-54

Query Match 53.6%; Score 75; DB 1; Length 1141;
Best Local Similarity 65.4%; Pred. No. 0.014;
Matches 17; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWLALRALGRPLPTS 26
Db 586 ARGDFAAQAQQLWLALRALGRPLPTS 611

RESULT 6
US-08-668-123-54
; Sequence 54, Application US/08668123
; Patent No. 5891631
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
/ COMPUTER READABLE FORM: disk
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/668,123
/ FILING DATE: 14-JUN-1996
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ PRIOR APPLICATION NUMBER: 08/131,365
/ FILING DATE: 01-OCT-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Parker, David L.
/ REGISTRATION NUMBER: 32,165
/ REFERENCE/DOCKET NUMBER: UTSD:372/PAR
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 418-3000

/
/
/ ZIP: 77210
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/252,991A
/ FILING DATE: 01-OCT-1999
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ PRIOR APPLICATION NUMBER: US/09/252,991A
/ FILING DATE: 1999-02-18
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Rubenfield et al.
/ REGISTRATION NUMBER: 32,165
/ REFERENCE/DOCKET NUMBER: UTSD:372/PAR
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (512) 474-7577
/ TELEFAX: (512) 474-7577
/ INFORMATION FOR SEQ ID NO: 54:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1141
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-252-991A-18345

Query Match 34.6%; Score 48.5; DB 4; Length 503;
Best Local Similarity 48.0%; Pred. No. 45;
Matches 12; Conservative 4; Mismatches 6; Indels 3; Gaps 1;

QY 1 ARGDFAAQAQQL--WLALRALGRP 22
Db 362 ADGDDAQAQVRVQWLARKGITAP 386

RESULT 7
US-09-252-991A-18345
; Sequence 18345, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18345
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
/ US-09-252-991A-18345

Query Match 34.6%; Score 48.5; DB 4; Length 503;
Best Local Similarity 48.0%; Pred. No. 45;
Matches 12; Conservative 4; Mismatches 6; Indels 3; Gaps 1;

QY 1 ARGDFAAQAQQL--WLALRALGRP 22
Db 362 ADGDDAQAQVRVQWLARKGITAP 386

RESULT 8
US-09-252-991A-25034
; Sequence 25034, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25034
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
/ US-09-252-991A-25034

Query Match 34.3%; Score 48; DB 4; Length 478;
```

```
Best Local Similarity 36.0%; Pred. No. 51;
Matches 9; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

Qy 1 ARGDFAQAQQLWLALRALGRPLPT 25
Db 219 SRSANTAPAREWNCSSRGCSPLPT 243

RESULT 9
US-09-252-991A-17953
; Sequence 17953, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; EARLIER FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 233
; SEQ ID NO 17953
; LENGTH: 849
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17953

Query Match 34.3%; Score 48; DB 4; Length 849;
Best Local Similarity 46.7%; Pred. No. 96;
Matches 14; Conservative 0; Mismatches 8; Indels 8; Gaps 1;

Qy 6 AQAQQL-----WLALRALGRPLPTSH 27
Db 657 ASASQLVRSEEREMWEALRALRKLAEH 686

RESULT 10
US-09-382-552-233
; Sequence 233, Application US/09382552
; Patent No. 6673909
; GENERAL INFORMATION:
; APPLICANT: Brown, Jr., Robert H.
; APPLICANT: Liu, Jing
; APPLICANT: Aoki, Masashi
; APPLICANT: Ho, Meng
; APPLICANT: Matsuda-Asada, Chie
; TITLE OF INVENTION: DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY AND LIMB
; FILE REFERENCE: 00786/39002
; CURRENT APPLICATION NUMBER: US/09/382,552
; PRIOR FILING DATE: 1999-08-25
; EARLIER FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 233
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 233
; LENGTH: 988
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-382-552-233

Query Match 34.3%; Score 48; DB 4; Length 988;
Best Local Similarity 47.8%; Pred. No. 11e+02;
Matches 11; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

Qy 4 DFAQAQQLWLAL--RALGRPLP 24
Db 685 DIEQKQMWVDLFPKALGRPG 707

RESULT 11
US-09-382-552-2
; Sequence 2, Application US/09382552
; Patent No. 6673909
; GENERAL INFORMATION:
; APPLICANT: Brown, Jr., Robert H.
; APPLICANT: Liu, Jing
; APPLICANT: Aoki, Masashi
; APPLICANT: Ho, Meng
; APPLICANT: Matsuda-Asada, Chie
; TITLE OF INVENTION: DYSFERLIN, A GENE MUTATED IN DISTAL MYOPATHY AND LIMB
; FILE REFERENCE: 00786/39002
; CURRENT APPLICATION NUMBER: US/09/382,552
; PRIOR FILING DATE: 1999-08-25
; EARLIER FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 233
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2080
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-382-552-2

Query Match 34.3%; Score 48; DB 4; Length 2080;
Best Local Similarity 47.8%; Pred. No. 2.6e+02;
Matches 11; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

Qy 4 DFAQAQQLWLAL--RALGRPLP 24
Db 1777 DIEQKQMWVDLFPKALGRPG 1799

RESULT 12
US-09-252-991A-31885
; Sequence 31885, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31885
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31885

Query Match 33.6%; Score 47; DB 4; Length 501;
Best Local Similarity 50.0%; Pred. No. 75;
Matches 10; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 8 AAQQLWLALRALGRPLPTSH 27
Db 80 AAHQLFALGFLGHPVEQAH 99

RESULT 13
US-08-466-390-2
; Sequence 2, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
; APPLICANT: TOURKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
```

```

; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 639 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-466-390-2

```

```

Query Match 33.6%; Score 47; DB 1; Length 639;
Best Local Similarity 47.8%; Pred. No. 98;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 6 AQAACQLWLALRGLRPLPTS 26
|:|||||:|
Db 444 ARKAQHLWLSVEALKYSMTS 464

```

```

RESULT 14
US-08-470-950-2
; Sequence 2, Application US/08470950
; Patent No. 5698439
; GENERAL INFORMATION:
; APPLICANT: TOKUATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:

```

```

; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 639 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-470-950-2
; Query Match 33.6%; Score 47; DB 1; Length 639;
; Best Local Similarity 47.6%; Pred. No. 98;
; Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
QY 6 AQAACQLWLALRGLRPLPTS 26
|:|||||:|
Db 444 ARKAQHLWLSVEALKYSMTS 464

```

```

RESULT 15
US-08-467-781-2
; Sequence 2, Application US/08467781
; Patent No. 5780596
; GENERAL INFORMATION:
; APPLICANT: TOKUATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,781
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 639 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-467-781-2

```

```

Query Match 33.6%; Score 47; DB 1; Length 639;
Best Local Similarity 47.6%; Pred. No. 98;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

```

```

QY 6 AQAACQLWLALRGLRPLPTS 26
|:|||||:|
Db 444 ARKAQHLWLSVEALKYSMTS 464

```

```

Search completed: November 10, 2004, 12:32:33
Job time : 9.04167 secs

```

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 15:53:52 ; Search time 26.0417 Seconds
(without alignments)
366.225 Million cell updates/sec

Title: US-10-092-750-22

Perfect score: 140

Sequence: 1 ARGDFAAQAQQLMLALRALGRPLPTSH 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/PCTUS_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	140	100.0	27	14	US-10-092-750-22
2	140	100.0	766	15	US-10-108-260A-3088
3	140	100.0	1147	15	US-10-429-160-6
4	140	100.0	1198	15	US-10-425-114-72712
5	74	52.9	1141	16	US-10-741-601-557
6	74	52.9	1141	16	US-10-741-601-558
7	65	46.4	92	14	US-10-029-386-28054
8	54.5	38.9	468	9	US-09-768-826-40
9	54.5	38.9	468	11	US-09-833-245-2260
10	54.5	38.9	551	9	US-09-897-214-8
11	54.5	38.9	565	9	US-09-768-826-58
12	54.5	38.9	566	14	US-10-093-463-190
13	54.5	38.9	647	14	US-10-093-463-188

14	54.5	38.9	778	14	US-10-093-463-192	Sequence 192, App
15	51	36.4	78	17	US-10-425-115-356053	Sequence 356053,
16	50.5	36.1	385	16	US-10-437-963-174932	Sequence 174932,
17	50.5	36.1	884	16	US-10-437-963-108154	Sequence 108154,
18	50	35.7	158	15	US-10-425-114-41739	Sequence 41739, A
19	50	35.7	172	16	US-10-437-963-104413	Sequence 104413,
20	50	35.7	177	17	US-10-425-115-366972	Sequence 366972,
21	50	35.7	212	15	US-10-425-114-39114	Sequence 39114, A
22	50	35.7	212	17	US-10-425-115-313297	Sequence 313297,
23	50	35.7	334	15	US-10-425-114-54215	Sequence 54215, A
24	50	35.7	385	15	US-10-425-114-49178	Sequence 49178, A
25	50	35.7	431	17	US-10-425-115-313295	Sequence 313295,
26	49.5	35.4	119	11	US-09-833-245-1187	Sequence 1187, Ap
27	49	35.0	61	17	US-10-425-115-307480	Sequence 307480,
28	49	35.0	134	16	US-10-437-963-138612	Sequence 138612,
29	49	35.0	181	16	US-10-437-963-167918	Sequence 167918,
30	49	35.0	1095	14	US-10-369-493-4059	Sequence 4059, Ap
31	48.5	34.6	103	16	US-10-437-963-119278	Sequence 119278,
32	48.5	34.6	476	15	US-10-282-122A-43433	Sequence 43433, A
33	48.5	34.6	1389	15	US-10-282-122A-70151	Sequence 70151, A
34	48	34.3	111	17	US-10-425-115-250826	Sequence 250826,
35	48	34.3	177	16	US-10-437-963-150984	Sequence 150984,
36	48	34.3	234	15	US-10-425-114-53833	Sequence 53833, A
37	48	34.3	249	17	US-10-425-115-214213	Sequence 214213,
38	48	34.3	299	15	US-10-424-599-258037	Sequence 258037,
39	48	34.3	322	14	US-10-314-657-69	Sequence 69, Appl
40	48	34.3	330	15	US-10-424-599-150210	Sequence 150210,
41	48	34.3	540	15	US-10-354-437-54	Sequence 54, Appl
42	48	34.3	712	9	US-09-815-242-11915	Sequence 11915, A
43	48	34.3	712	15	US-10-282-122A-66476	Sequence 66476, A
44	48	34.3	2080	10	US-09-382-860-2	Sequence 2, Appl
45	48	34.3	2080	16	US-10-408-765A-411	Sequence 411, App

ALIGNMENTS

RESULT 1
US-10-092-750-22
; Sequence 22, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; PRIOR FILING DATE: 2002-03-07
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 27
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-22

Query Match	100.0%;	Score 140;	DB 14;	Length 27;
Best Local Similarity	100.0%;	Pred No. 1.1e-12;		
Matches	27;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	ARGDFAAQAQQLMLALRALGRPLPTSH	27	
Db	1	ARGDFAAQAQQLMLALRALGRPLPTSH	27	
RESULT 2				
US-10-108-260A-3088				
; Sequence 3088, Application US/10108260A				
; Publication No. US20040005560A1				
; GENERAL INFORMATION:				

; APPLICANT: HELIX RESEARCH INSTITUTE
 ; TITLE OF INVENTION: No. US20040005560A1el full length cDNA
 ; FILE REFERENCE: HI-A0106
 ; CURRENT APPLICATION NUMBER: US/10/108,260A
 ; CURRENT FILING DATE: 2002-03-27
 ; NUMBER OF SEQ ID NOS: 5458
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 3088
 ; LENGTH: 766
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-108-260A-3088

Query Match 100.0%; Score 140; DB 15; Length 766;
 Best Local Similarity 100.0%; Pred. No. 2.9e-11;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWLALRGRPLPTSH 27
 DB 213 ARGDFAAQAQQLWLALRGRPLPTSH 239

RESULT 3
 US-10-429-160-6
 ; Sequence 6, Application US/10429160
 ; Publication No. US2004003276A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Ward, Teresa R
 ; APPLICANT: Mao, Mao
 ; APPLICANT: Linsley, Peter S
 ; APPLICANT: Lund, Lund
 ; TITLE OF INVENTION: LXR Ligand Induced Genes and Proteins
 ; FILE REFERENCE: RS0200
 ; CURRENT APPLICATION NUMBER: US/10/429,160
 ; CURRENT FILING DATE: 2003-05-02
 ; PRIOR APPLICATION NUMBER: US 60/377,714
 ; PRIOR FILING DATE: 2002-05-03
 ; NUMBER OF SEQ ID NOS: 111
 ; SOFTWARE: Patentin version 3.1
 ; SEQ ID NO 6
 ; LENGTH: 1147
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-429-160-6

Query Match 100.0%; Score 140; DB 15; Length 1147;
 Best Local Similarity 100.0%; Pred. No. 4.4e-11;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWLALRGRPLPTSH 27
 DB 594 ARGDFAAQAQQLWLALRGRPLPTSH 620

RESULT 4
 US-10-425-114-72712
 ; Sequence 72712, Application US/10425114
 ; Publication No. US20040034888A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Liu, Jingdong
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Screen, Steven E.
 ; APPLICANT: Tabaska, Jack E.
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53313)B
 ; CURRENT APPLICATION NUMBER: US/10/425,114
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 73128
 ; SEQ ID NO 72712
 ; LENGTH: 1198

; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: LIB4750-126-A8_FLI pep
 US-10-425-114-72712

Query Match 100.0%; Score 140; DB 15; Length 1198;
 Best Local Similarity 100.0%; Pred. No. 4.6e-11;
 Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWLALRGRPLPTSH 27
 DB 645 ARGDFAAQAQQLWLALRGRPLPTSH 671

RESULT 5
 US-10-741-601-557
 ; Sequence 557, Application US/10741601
 ; Publication No. US20040166519A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; FILE REFERENCE: CL001500
 ; CURRENT APPLICATION NUMBER: US/10/741,601
 ; CURRENT FILING DATE: 2003-12-22
 ; NUMBER OF SEQ ID NOS: 26415
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 557
 ; LENGTH: 1141
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-741-601-557

Query Match 52.9%; Score 74; DB 16; Length 1141;
 Best Local Similarity 65.4%; Pred. No. 0.11;
 Matches 17; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWLALRGRPLPTS 26
 DB 586 ARGDFAAAGNLOTCLAVLGRALPTS 611

RESULT 6
 US-10-741-601-558
 ; Sequence 558, Application US/10741601
 ; Publication No. US20040166519A1
 ; GENERAL INFORMATION:
 ; APPLICANT: CARGILL, Michele et al.
 ; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
 ; FILE REFERENCE: CL001500
 ; CURRENT APPLICATION NUMBER: US/10/741,601
 ; CURRENT FILING DATE: 2003-12-22
 ; NUMBER OF SEQ ID NOS: 26415
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 558
 ; LENGTH: 1141
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-741-601-558

Query Match 52.9%; Score 74; DB 16; Length 1141;
 Best Local Similarity 65.4%; Pred. No. 0.11;
 Matches 17; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWLALRGRPLPTS 26
 DB 586 ARGDFAAAGNLOTCLAVLGRALPTS 611

RESULT 7
 US-10-029-386-28054

```

; Sequence 28054, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 28054
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO CHR22 22.0
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
; OTHER INFORMATION: SWISSPROT HIT: Q12772, EVALUATE 2.00e-37
US-10-029-386-28054

Query Match      46.4%; Score 65; DB 14; Length 92;
Best Local Similarity 62.5%; Pred. No. 0.17;
Matches 15; Conservative 0; Mismatches 9; Indels 9; Gaps 0;

QY 3 GDFAAQAQQLWLALRGRLPLTS 26
DB 1 GDFAAAGNLTCLAVIGRALPTS 24

RESULT 8
US-09-768-826-40
; Sequence 40, Application US/09768826
; Patent No. US20020012566A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: 18 human secreted proteins
; FILE REFERENCE: PF512P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759
; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 40
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-768-826-40

Query Match      38.9%; Score 54.5; DB 9; Length 468;
Best Local Similarity 37.5%; Pred. No. 26;
Matches 15; Conservative 3; Mismatches 7; Indels 15; Gaps 2;

QY 3 GDF-----AQAQQLW-----LALRALGRPLPTSH 27
DB 107 GFSCPEPLIARHTQRLWLVLEGORATLRCRALGDPAPTMH 146

RESULT 9
US-09-833-245-2260
; Sequence 2260, Application US/09833245
; Publication No. US20040010134A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Albumin Fusion Proteins
; FILE REFERENCE: PF546PCI
; CURRENT APPLICATION NUMBER: US/09/833,245
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: 60/229,358
; PRIOR FILING DATE: 2000-04-12
; PRIOR APPLICATION NUMBER: 60/256,931
; PRIOR FILING DATE: 2000-12-21
; PRIOR APPLICATION NUMBER: 60/199,384
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 2267
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2260
; LENGTH: 468
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-833-245-2260

Query Match      38.9%; Score 54.5; DB 11; Length 468;
Best Local Similarity 37.5%; Pred. No. 26;
Matches 15; Conservative 3; Mismatches 7; Indels 15; Gaps 2;

QY 3 GDF-----AQAQQLW-----LALRALGRPLPTSH 27
DB 107 GFSCPEPLIARHTQRLWLVLEGORATLRCRALGDPAPTMH 146

RESULT 10
US-09-897-214-8
; Sequence 8, Application US/09897214
; Patent No. US20020076779A1
; GENERAL INFORMATION:
; APPLICANT: Thayer, Edward C.
; APPLICANT: Sheppard, Paul O.
; APPLICANT: Presnell, Scott R.
; TITLE OF INVENTION: Leucine-Rich Repeat Proteins, Zlrr7,
; TITLE OF INVENTION: Zlrr8, and Zlrr9
; FILE REFERENCE: 01-27
; CURRENT APPLICATION NUMBER: US/09/897,214
; CURRENT FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: 60/215,446
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-897-214-8

Query Match      38.9%; Score 54.5; DB 9; Length 551;
Best Local Similarity 37.5%; Pred. No. 31;
Matches 15; Conservative 3; Mismatches 7; Indels 15; Gaps 2;

QY 3 GDF-----AQAQQLW-----LALRALGRPLPTSH 27
DB 274 GFSCPEPLIARHTQRLWLVLEGORATLRCRALGDPAPTMH 313

RESULT 11
US-09-768-826-58
; Sequence 58, Application US/09768826
; Patent No. US20020012966A1
; GENERAL INFORMATION:
; APPLICANT: Shi et al.
; TITLE OF INVENTION: 18 human secreted proteins
; FILE REFERENCE: PF512P1
; CURRENT APPLICATION NUMBER: US/09/768,826
; CURRENT FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: PCT/US00/22350
; PRIOR FILING DATE: 2000-08-15
; PRIOR APPLICATION NUMBER: 60/148,759

```

```

; PRIOR FILING DATE: 1999-08-16
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 58
; LENGTH: 565
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (270)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-768-526-58

Query Match      38.9%; Score 54.5; DB 9; Length 565;
Best Local Similarity 37.5%; Pred. No. 31;
Matches 15; Conservative 3; Mismatches 7; Indels 15; Gaps 2;

Qy 3 GDF-----AQAQQQLW-----LALRALGRPLPTSH 27
   | | | | | | | | | | | | | | | | | | | | | |
Db 204 GFSCPEPLIARHTORLWLVLEGGQRATLRCRAGDPAFTMH 243

RESULT 12
US-10-093-463-190
; Sequence 190, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glenna
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Anderson, David
; APPLICANT: Patturajan, Meera
; APPLICANT: Miller, Charles
; APPLICANT: Taupier, Raymond J. Jr.
; TITLE OF INVENTION: No. US20030208039A1 Antibodies that Bind to Antigenic Polypepti
; FILE REFERENCE: 21402-230A (Cura 590AT)
; CURRENT APPLICATION NUMBER: US/10/093,463
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 60/283,675
; PRIOR FILING DATE: 2001-04-14
; PRIOR APPLICATION NUMBER: 60/338,092
; PRIOR FILING DATE: 2001-12-03
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,101
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/325,681
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: 60/304,354

```

```

; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/279,995
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/294,899
; PRIOR FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: 60/287,424
; PRIOR FILING DATE: 2001-04-30
; PRIOR APPLICATION NUMBER: 60/299,027
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/309,198
; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/281,194
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/274,194
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/330,380
; PRIOR FILING DATE: 2001-10-18
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/288,342
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/275,578
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 190
; LENGTH: 566
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-093-463-190

Query Match      38.9%; Score 54.5; DB 14; Length 566;
Best Local Similarity 37.5%; Pred. No. 32;
Matches 15; Conservative 3; Mismatches 7; Indels 15; Gaps 2;

Qy 3 GDF-----AQAQQQLW-----LALRALGRPLPTSH 27
   | | | | | | | | | | | | | | | | | | | | | |
Db 274 GFSCPEPLIARHTORLWLVLEGGQRATLRCRAGDPAFTMH 313

RESULT 13
US-10-093-463-188
; Sequence 188, Application US/10093463
; Publication No. US20030208039A1
; GENERAL INFORMATION:
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Shenoy, Suresh
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Gusev, Vladimir
; APPLICANT: Pochart, Pascal
; APPLICANT: Zhong, Mei
; APPLICANT: Rastelli, Luca
; APPLICANT: Mezes, Peter
; APPLICANT: Smithson, Glenna
; APPLICANT: Guo, Xiaojia
; APPLICANT: Gerlach, Valerie
; APPLICANT: Casman, Stacie
; APPLICANT: Boldog, Ferenc
; APPLICANT: Li, Li
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Tchernev, Velizar
; APPLICANT: Gangolli, Esha
; APPLICANT: Vernet, Corine
; APPLICANT: Pena, Carol
; APPLICANT: Burgess, Catherine
; APPLICANT: Liu, Xiaohong
; APPLICANT: Spytek, Kimberly
; APPLICANT: Gorman, Linda
; APPLICANT: Spaderna, Steven
; APPLICANT: Voss, Edward
; APPLICANT: Malyankar, Uriel

```



```

1  APPLICANT: Rastelli, Luca
2  APPLICANT: Mezes, Peter
3  APPLICANT: Smithson, Glennnda
4  APPLICANT: Guo, Xiaojia
5  APPLICANT: Gerlach, Valerie
6  APPLICANT: Casman, Stacie
7  APPLICANT: Boldog, Ferenc
8  APPLICANT: Li, Li
9  APPLICANT: Zerhusen, Bryan
10 APPLICANT: Tchernev, Velizar
11 APPLICANT: Gangolli, Esha
12 APPLICANT: Vernet, Corine
13 APPLICANT: Pena, Carol
14 APPLICANT: Burgess, Catherine
15 APPLICANT: Liu, Xiaohong
16 APPLICANT: Spytek, Kimberly
17 APPLICANT: Gorman, Linda
18 APPLICANT: Spaderna, Steven
19 APPLICANT: Voss, Edward
20 APPLICANT: Malyankar, Uriel
21 APPLICANT: Anderson, David
22 APPLICANT: Patturajan, Meera
23 APPLICANT: Miller, Charles
24 APPLICANT: Taupier, Raymond J. Jr.
25  TITLE OF INVENTION: NO. US20030208039a1el Antibodies that Bind to Antigenic Polypepti
26  TITLE OF INVENTION: Encoding The Antigens, and Methods of Use.
27  FILE REFERENCE: 21402-290A (Cura 590AT)
28  CURRENT APPLICATION NUMBER: US/10/093,463
29  CURRENT FILING DATE: 2002-06-24
30  PRIOR APPLICATION NUMBER: 60/283,675
31  PRIOR FILING DATE: 2001-04-14
32  PRIOR APPLICATION NUMBER: 60/338,092
33  PRIOR FILING DATE: 2001-12-03
34  PRIOR APPLICATION NUMBER: 60/274,281
35  PRIOR FILING DATE: 2001-03-08
36  PRIOR APPLICATION NUMBER: 60/274,101
37  PRIOR FILING DATE: 2001-03-08
38  PRIOR APPLICATION NUMBER: 60/325,681
39  PRIOR FILING DATE: 2001-09-27
40  PRIOR APPLICATION NUMBER: 60/304,354
41  PRIOR FILING DATE: 2001-07-10
42  PRIOR APPLICATION NUMBER: 60/279,995
43  PRIOR FILING DATE: 2001-03-30
44  PRIOR APPLICATION NUMBER: 60/294,899
45  PRIOR FILING DATE: 2001-05-31
46  PRIOR APPLICATION NUMBER: 60/287,424
47  PRIOR FILING DATE: 2001-04-30
48  PRIOR APPLICATION NUMBER: 60/299,027
49  PRIOR FILING DATE: 2001-06-18
50  PRIOR APPLICATION NUMBER: 60/309,198
51  PRIOR FILING DATE: 2001-07-31
52  PRIOR APPLICATION NUMBER: 60/281,194
53  PRIOR FILING DATE: 2001-04-04
54  PRIOR APPLICATION NUMBER: 60/274,194
55  PRIOR FILING DATE: 2001-03-08
56  PRIOR APPLICATION NUMBER: 60/274,849
57  PRIOR FILING DATE: 2001-03-09
58  PRIOR APPLICATION NUMBER: 60/330,380
59  PRIOR FILING DATE: 2001-10-18
60  PRIOR APPLICATION NUMBER: 60/275,235
61  PRIOR FILING DATE: 2001-03-12
62  PRIOR APPLICATION NUMBER: 60/288,342
63  PRIOR FILING DATE: 2001-05-03
64  PRIOR APPLICATION NUMBER: 60/275,578
65  PRIOR FILING DATE: 2001-03-13
66  NUMBER OF SEQ ID NOS: 370
67  SOFTWARE: PatentIn Ver. 2.1
68  SEQ ID NO 192
69  LENGTH: 778
70  TYPE: PRT
71  ORGANISM: Homo sapiens
72  US-10-093-463-192

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:36:51 ; Search time 5.45833 Seconds
(without alignments)
475.942 Million cell updates/sec

Title: US-10-092-750-22
Perfect score: 140
Sequence: 1 ARGDFAAQAQQLWLALRALGRPLPTSH 27

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 791*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	129	92.1	927	2 A48085	transcription fact
2	129	92.1	1133	2 A54164	sterol regulatory
3	75	53.6	1139	2 B54962	sterol regulatory
4	75	53.6	1141	2 A54962	sterol regulatory
5	51	36.4	352	2 T10305	adenylate cyclase
6	50.5	36.1	904	2 A84212	hypothetical prote
7	49	35.0	410	2 D75475	lycopen cyclase -
8	48.5	34.6	72	2 A36379	lymphotoxin Cl pre
9	48.5	34.6	476	2 H83265	probable D-alanyl-
10	48	34.3	347	2 T34919	oxidoreductase - S
11	48	34.3	503	2 T46852	anthranilate synth
12	48	34.3	712	2 E83226	ATP-dependent DNA
13	47.5	33.9	109	2 C64386	hypothetical prote
14	47.5	33.9	336	2 E84295	hypothetical prote
15	47.5	33.9	444	2 F75387	NADH dehydrogenase
16	47.5	33.9	3149	1 Q0B88	BPLF1 protein - hu
17	47	33.6	312	2 E87350	acetyltransferase,
18	47	33.6	325	2 D87659	Sua5/YciO/YrdC fam
19	47	33.6	757	2 C32052	FixII protein - Rhi
20	47	33.6	757	2 C95344	fixII copper trans
21	46.5	33.2	370	2 AE0289	conserved hypothet
22	46.5	33.2	586	1 JVBEP3	DNA maturase B - p
23	46.5	33.2	1067	2 T06672	probable aminocacyl
24	46.5	33.2	1067	2 T48850	glycine-tRNA ligase
25	46	32.9	383	2 T05590	hypothetical prote
26	46	32.9	428	2 AC2986	sugar kinase limpo
27	46	32.9	460	2 D98297	probable rhamnose
28	46	32.9	709	2 S75212	comE protein - Syn
29	46	32.9	1402	2 F84480	probable retroelem

alkaline phosphata
hypothetical prote
GDPF family prote
hypothetical prote
hypothetical prote
cellobiose phospho
cellobiose phospho
hypothetical prote
conserved hypothet
hypothetical prote
conserved hypothet
hypothetical prote
neural cell adhesi
hypothetical prote
phage lysis protei
chpB protein - pse
Ra-reactive factor

ALIGNMENTS

RESULT 1

A48085
transcription factor ADD1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-May-1994 #sequence_revision 04-Sep-1999
C:Accession: A48085
R:Tontoz, P.; Kim, J.B.; Graves, R.A.; Spiegelman, B.M.
Mol. Cell. Biol. 13, 4753-4759, 1993
A:Title: ADD1: a novel helix-loop-helix transcription factor associated with adipocyte de
A:Reference number: A48085; MUID:93330269; PMID:8336713
A:Accession: A48085
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-927 <TON>
A:Cross-references: GB:L16995
A:Superfamily: sterol regulatory element binding protein
C:Keywords: leucine zipper

Query Match 92.1%; Score 129; DB 2; Length 927;
Best Local Similarity 92.6%; Pred. No. 1.1e-10;
Matches 25; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWLALRALGRPLPTSH 27
DB 559 ARGDFAAQAQQLWLALQALGRPLPTSN 585

RESULT 2

A54164
sterol regulatory element-binding protein 1 - Chinese hamster
C:Species: Crictetus griseus (Chinese hamster)
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 09-Jul-2004
C:Accession: A54164
R:Sato, R.; Yang, J.; Wang, X.; Evans, M.J.; Ho, Y.K.; Goldstein, J.L.; Brown, M.S.
J. Biol. Chem. 269, 17267-17273, 1994
A:Title: Assignment of the membrane attachment, DNA binding, and transcriptional activat
A:Reference number: A54164; MUID:94274723; PMID:8006035
A:Accession: A54164
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1133 <SAT>
A:Cross-references: UNIPROT:Q60416; GB:U09103; NID:9516002; PIDN:AAA20085.1; PID:9516003
C:Superfamily: sterol regulatory element binding protein

Query Match 92.1%; Score 129; DB 2; Length 1133;
Best Local Similarity 92.6%; Pred. No. 1.4e-10;
Matches 25; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWLALRALGRPLPTSH 27
DB 582 ARGDFAAQAQQLWLALQALGRPLPTSN 608

RESULT 3

B54962

sterol regulatory element binding protein 2 precursor - Chinese hamster

C:Species: Cricetulus griseus (Chinese hamster)

C:Date: 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 09-Jul-2004

C:Accession: B54962

R:Yang, J.; Sato, R.; Goldstein, J.L.; Brown, M.S.

Genes Dev. 8, 1910-1919, 1994

A:Title: Sterol-resistant transcription in CHO cells caused by gene rearrangement that

A:Reference number: A54962; MUID:95047343; PMID:7958866

A:Accession: B54962

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-1139 <YAN>

A:Cross-references: UNIPROT:Q60429; GB:U12330

A:Note: 493 Ser was also found

C:Superfamily: sterol regulatory element binding protein

C:Keywords: DNA binding; membrane protein

Query Match 53.6%; Score 75; DB 2; Length 1139;
 Best Local Similarity 65.4%; Pred. No. 0.0097;
 Matches 17; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ARGDFAAQOLWLALRGLRPLTS 26

DB 584 ARGDFAAAAANLQTCLSLVGLRPLTS 609

RESULT 4

A54962

sterol regulatory element binding protein 2 precursor - human

N:Alternate names: basic-helix-loop-helix-leucine zipper transcription factor SREBP-2

C:Species: Homo sapiens (man)

C:Date: 23-Mar-1995 #sequence_revision 05-Apr-1995 #text_change 09-Jul-2004

C:Accession: A49397; A54962

R:Hua, X.; Yokoyama, C.; Wu, J.; Briggs, M.R.; Brown, M.S.; Goldstein, J.L.; Wang, X.

Proc. Natl. Acad. Sci. U.S.A. 90, 11603-11607, 1993

A:Title: SREBP-2, a second basic-helix-loop-helix-leucine zipper protein that stimulates

A:Reference number: A49397; MUID:94089681; PMID:7903453

A:Accession: A49397

A:Molecule type: mRNA

A:Residues: 1-1141 <HUA>

A:Cross-references: UNIPROT:Q12772; GB:U02031; NID:G451329; PIDN:AAA50746.1; PID:G451330

A:Experimental source: HeLa cells

A:Note: sequence extracted from NCBI backbones (NCBI:P:140987)

R:Yang, J.; Sato, R.; Goldstein, J.L.; Brown, M.S.

Genes Dev. 8, 1910-1919, 1994

A:Title: Sterol-resistant transcription in CHO cells caused by gene rearrangement that

A:Reference number: A54962; MUID:95047343; PMID:7958866

A:Contents: annotation

C:Comment: Under conditions of sterol depletion, this membrane-bound transcription facto

C:Genetics:

A:Gene: GDB:SREBP2; SREBP2

A:Cross-references: GDB:273625

A:Map position: Xq24-Xqter

C:Superfamily: sterol regulatory element binding protein

C:Keywords: DNA binding; membrane protein; transcription factor

Query Match 53.6%; Score 75; DB 2; Length 1141;
 Best Local Similarity 65.4%; Pred. No. 0.0098;
 Matches 17; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ARGDFAAQOLWLALRGLRPLTS 26

DB 586 ARGDFAAAAANLQTCLAVLGRPLTS 611

RESULT 5

T10905

adenylate cyclase (EC 4.6.1.1) - Stigmatella aurantiaca

C:Species: Stigmatella aurantiaca

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004

C:Accession: T10905

R:Coudart-Cavalli, M.P.; Sismeiro, O.; Danchin, A.

submitted to the EMBL Data Library, January 1998

A:Description: Bifunctional structure of two adenylyl cyclases from the myxobacterium St;

A:Reference number: Z17211

A:Accession: T10905

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-352 <COU>

A:Cross-references: UNIPROT:P40138; EMBL:AJ223795

C:Genetics:

A:Gene: cyaB

C:Function:

A:Description: involved in cAMP synthesis

A:Superfamily: sensory transduction system regulatory protein; response regulator homolog

C:Keywords: phosphorus-oxygen lyase

Query Match 36.4%; Score 51; DB 2; Length 352;
 Best Local Similarity 55.0%; Pred. No. 9.3;
 Matches 11; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

QY 4 DPAQAAQOLWLALRGLRPL 23

DB 247 ELRQAAQAEWAANERLGRPL 266

RESULT 6

A84212

hypothetical protein Vng0537c [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004

C:Accession: A84212

R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.

; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablor

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li;

A:Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: A84160; MUID:20504483; PMID:11016950

A:Accession: A84212

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-904 <STO>

A:Cross-references: UNIPROT:Q9HRU7; GB:AE004437; NID:G10580137; PIDN:AAG19061.1; GSPDB:GN

C:Genetics:

A:Gene: VNG0537C

Query Match 36.1%; Score 50.5; DB 2; Length 904;
 Best Local Similarity 54.5%; Pred. No. 28;
 Matches 12; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

QY 4 DPAQAAQOLWLALRGLRPLPT 25

DB 165 DPAEIRLYLGLTA-GRPLPT 185

RESULT 7

D75475

lycopen cyclase - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C:Accession: D75475

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; F

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Mat

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896; PMID:1056266

A:Accession: D75475

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-410 <WHI>

A:Cross-references: UNIPROT:Q9RW68; GB:AE001934; GB:AE000513; NID:G6458504; PIDN:AAF1037:


```

A:Experimental source: strain PA01
C:Genetics:
A:Gene: recQ; PA3344
C:Superfamily: recQ protein; recQ helicase homology

Query Match      34.3%  Score 48;  DB 2;  Length 712;
Best Local Similarity 46.7%  Pred. No. 51;
Matches 14;  Conservative 0;  Mismatches 8;  Indels 8;  Gaps 1;

QY 6 AQAQQQL-----WIALRALGRPLPTSH 27
    |||||
DB 520 ASAASQVRSEREMWELRALRRKLAEEH 549

RESULT 13
C64386
Hypothetical protein MJ0691 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 09-Jul-2004
C:Accession: C64386
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
rson, J.D.; Sadov, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
A:Reference number: A64300; MUID:96337999; PMID:8688087
A:Accession: C64386
A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-109 <BUL>
A:Cross-references: UNIPROT:Q58103; GB:U67516; GB:L77117; NID:92826307; PIDN:AA898686.1;
C:Genetics:
A:Map position: REV618302-617973
C:Superfamily: DNA-binding protein, TFIIR19 type

Query Match      33.9%  Score 47.5;  DB 2;  Length 109;
Best Local Similarity 44.4%  Pred. No. 9.4;
Matches 12;  Conservative 5;  Mismatches 9;  Indels 1;  Gaps 1;

QY 1 ARGDFAQAQQLWALRALGR-PLPTS 26
    |||||
DB 57 ARPEFAEVEVQLIQALGRPLPLS 83

RESULT 14
E84295
Hypothetical protein Vng1410h [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: E84295
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
; Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablo
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1
A:Reference number: A84160; MUID:20504483; PMID:11016950
A:Accession: E84295
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-336 <STO>
A:Cross-references: UNIPROT:Q9HPY8; GB:AE004437; NID:g10580915; PIDN:AAG19729.1; GSPDB:G
C:Genetics:
A:Gene: Vng1410H

Query Match      33.9%  Score 47.5;  DB 2;  Length 336;
Best Local Similarity 50.0%  Pred. No. 29;
Matches 13;  Conservative 1;  Mismatches 11;  Indels 1;  Gaps 1;

QY 1 ARGDFAQAQQLWALRALGRPLPTS 26
    |||||
DB 194 AAADALAEALQNWELRALG-PEPAS 218

RESULT 15
F75387
NADH dehydrogenase I, F subunit - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: F75387
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; F
; M.; Shen, M.; Vamathevan, J.U.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896; PMID:10567266
A:Accession: F75387
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-444 <WHI>
A:Cross-references: UNIPROT:Q9RU92; GB:AE001994; GB:AE000513; NID:96459259; PIDN:AAF1106;
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1500
A:Map position: 1
C:Superfamily: NADH dehydrogenase (ubiquinone) chain F; NADH dehydrogenase (ubiquinone)
Query Match      33.9%  Score 47.5;  DB 2;  Length 444;
Best Local Similarity 45.5%  Pred. No. 38;
Matches 10;  Conservative 5;  Mismatches 4;  Indels 3;  Gaps 1;

QY 2 RGDFQAQAQQLWAL---RALG 20
    |||||
DB 139 RGEYVLAERLWAAIHARAAG 160

Search completed: November 10, 2004, 12:29:28
Job time : 7.45833 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:27:01 ; Search time 29.5417 Seconds
(without alignments)
525.871 Million cell updates/sec

Title: US-10-092-750-22

Perfect score: 140
Sequence: 1 ARGDFAAQAQQLWALRALGRPLPTSH 27

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt 02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	140	100.0	686	2 Q6PJ36	Q6PJ36 homo sapien
2	140	100.0	686	2 AAH23621	AAH23621 homo sapien
3	140	100.0	750	2 Q8TAK9	Q8TAK9 homo sapien
4	140	100.0	993	2 Q7Y802	Q7Y802 sus scrofa
5	140	100.0	1147	1 SREI HUMAN	SREI HUMAN
6	140	100.0	1147	2 Q6PFW7	Q6PFW7 homo sapien
7	140	100.0	1147	2 AAH57388	AAH57388 homo sapi
8	140	100.0	1151	2 Q6RIB9	Q6RIB9 sus scrofa
9	140	100.0	1151	2 AAS18238	AAS18238 sus scrofa
10	140	100.0	1177	2 Q6P4R7	Q6P4R7 homo sapien
11	140	100.0	1177	2 AAH63281	AAH63281 homo sapi
12	129	92.1	1024	1 SREI RAT	SREI RAT
13	129	92.1	1133	1 SREI_CRIGR	SREI_CRIGR
14	124	88.6	866	2 Q99UK7	Q99UK7 mus musculus
15	124	88.6	1134	1 SREI_MOUSE	SREI_MOUSE
16	124	88.6	1134	2 Q8C733	Q8C733 mus musculus
17	124	88.6	1134	2 AAH56922	AAH56922 mus muscu
18	96	68.6	1115	2 Q902M5	Q902M5 gallus gall
19	75	53.6	1139	1 SRE2_CRIGR	SRE2_CRIGR
20	75	53.6	1141	1 SRE2_HUMAN	SRE2_HUMAN
21	74	52.9	1141	2 AAH56158	AAH56158 homo sapi
22	74	52.9	1193	2 Q86V36	Q86V36 homo sapien
23	72	51.4	1125	2 Q925C2	Q925C2 mus musculus
24	57	40.7	232	2 Q8PWW2	Q8PWW2 xanthomonas
25	57	40.7	1088	2 Q6G026	Q6G026 xenopus lae
26	56.5	40.4	636	2 Q80XU8	Q80XU8 mus musculus
27	56.5	40.4	636	2 Q8K3C4	Q8K3C4 mus musculus
28	56.5	40.4	636	2 BAC38259	BAC38259 m 16 days
29	56	40.0	415	2 Q73RMI	Q73RMI treponema d
30	56	40.0	415	2 AAS10565	AAS10565 treponema
31	54.5	38.9	460	2 Q61PL6	Q61PL6 homo sapien

32	54.5	38.9	460	2 AAH71866	AAH71866 homo sapi
33	54.5	38.9	528	2 AAH27475	AAH27475 homo sapi
34	54.5	38.9	541	2 Q6PK41	Q6PK41 homo sapien
35	54.5	38.9	541	2 AAH07718	AAH07718 homo sapi
36	54.5	38.9	635	2 Q6PJ36	Q6PJ36 homo sapien
37	54.5	38.9	635	2 AAH15581	AAH15581 homo sapi
38	52	37.1	80	2 Q8DGR8	Q8DGR8 synchococcc
39	52	37.1	251	2 Q8PAXO	Q8PAXO xanthomonas
40	51.5	36.8	318	1 FMT_SHEON	FMT_SHEON
41	51	36.4	352	1 CVAB_STIAU	CVAB_STIAU
42	51	36.4	892	2 Q99PK4	Q99PK4 streptomyce
43	51	36.4	901	2 Q73YTO	Q73YTO mycobacteri
44	51	36.4	901	2 AAS04192	AAS04192 mycobacte
45	51	36.4	982	2 Q9VIG1	Q9VIG1 drosophila

ALIGNMENTS

RESULT 1
Q6PJ36 PRELIMINARY; PRT; 686 AA.
AC Q6PJ36;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE SREBFl protein (Fragment).
GN Name=SREBFl;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=22388257; PubMed=12477932;
RA Klausner R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Strausberg R.L., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stopleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
RA Farey J., Helton E., Kettelman M., Madan A., Rodriguez S., Shevchenko Y.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Small U., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RC TISSUE=Placenta;
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (Feb-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC023621; AAH23621.1; -.
FT NON TER 1
SQ SEQUENCE 686 AA; 74450 MW; 7439AF380530A6D2 CRC64;

Query Match 100.0%; Score 140; DB 2; Length 686;

Best Local Similarity 100.0%; Pred. No. 6.4e-11; Mismatches 0; Indels 0; Gaps 0;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWALRALGRPLPTSH 27

Db 133 ARGDFAAQAQQLWALRALGRPLPTSH 159

```

RESULT 2
AAH23621 PRELIMINARY; PRT; 686 AA.
ID AAH23621
AC AAH23621 (TREMBlrel. 27, Created)
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DE SREBF1 protein (Fragment).
DE SREBF1
GN SREBF1
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=2238257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heish F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC023621; AAH23621.1; -
FT NON_TER 1
SQ SEQUENCE 686 AA; 74450 MW; 7439AF380530A6D2 CRC64;

Query Match 100.0%; Score 140; DB 2; Length 686;
Best Local Similarity 100.0%; Pred. No. 6.4e-11;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWLALRALGRPLPTSH 27
DB 133 ARGDFAAQAQQLWLALRALGRPLPTSH 159

RESULT 3
Q8TAK9 PRELIMINARY; PRT; 750 AA.
ID Q8TAK9
AC Q8TAK9;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.

RESULT 4
Q7YS02 PRELIMINARY; PRT; 993 AA.
ID Q7YS02
AC Q7YS02;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Sterol regulatory binding transcription factor 1 (Fragment).
GN Name=SREBP-1c;
OS Sus scrofa (Pig)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
CX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RA Palin M.F., Beaudry D., Murphy B.D.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY307771; AAP74567.1; -
DR InterPro; IPR001092; HLH_Basic.
DR InterPro; IPR000566; Lipoclin_cytFABP.
DR Pfam; PF0010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00888; HLH; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 993 AA; 105340 MW; 7E6F9F9BD392843A CRC64;

Query Match 100.0%; Score 140; DB 2; Length 993;
Best Local Similarity 100.0%; Pred. No. 9.1e-11;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWLALRALGRPLPTSH 27
DB 436 ARGDFAAQAQQLWLALRALGRPLPTSH 462

RESULT 5
SRE1_HUMAN STANDARD; PRT; 1147 AA.
ID SRE1_HUMAN
AC P36956; Q16062;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DE 01-OCT-2004 (Rel. 45, Last annotation update)
DE Sterol regulatory element binding protein-1 (SREBP-1) (Sterol
DE regulatory element-binding transcription factor 1).
GN Name=SREBF1; Synonyms=SREBP1;
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND ALTERNATIVE SPLICING.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=94006541; PubMed=8402897;
RA Yokoyama C., Wang X., Briggs M.R., Admon A., Wu J., Hua X.,
RA Goldstein J.L., Brown M.S.;
RT "SREBP-1, a basic-helix-loop-helix-leucine zipper protein that
RT controls transcription of the low density lipoprotein receptor gene."

```


Cell 75:187-197(1993).
 [2]
 SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 TISSUE=Fetal brain;
 MEDLINE=95278936; PubMed=7759101;
 Hua X., Wu J., Goldstein J.L., Brown M.S., Hobbs H.H.;
 "Structure of the human gene encoding sterol regulatory element
 binding protein-1 (SREBF1) and localization of SREBF1 and SREBF2 to
 chromosomes 17p11.2 and 22q13.";
 Genomics 25:667-673(1995).
 [3]
 CHARACTERIZATION, AND MUTAGENESIS.
 MEDLINE=96215341; PubMed=8626610;
 Hua X., Sakai J., Brown M.S., Goldstein J.L.;
 "Regulated cleavage of sterol regulatory element binding proteins
 requires sequences on both sides of the endoplasmic reticulum
 membrane.";
 J. Biol. Chem. 271:10379-10384(1996).
 [4]
 X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 319-394.
 MEDLINE=98298442; PubMed=9634703;
 Parraga A., Bellolell L., Parre-D'Amare A.R., Burley S.K.;
 "Co-crystal structure of sterol regulatory element binding protein 1a
 at 2.3-A resolution.";
 Structure 6:661-672(1998).
 -1- FUNCTION: Transcriptional activator that binds to the sterol
 regulatory element 1 (SRE-1) (5'-ATCACCCAC-3'). Has dual sequence
 specificity, binding to both an E-box motif (5'-ATCACGTGA-3') and
 to SRE-1 (5'-ATCACCCAC-3'). Regulates the transcription of genes
 for sterol biosynthesis and the LDL receptor gene.
 -1- SUBUNIT: Forms a tight complex with SCAP in the ER membrane.
 Efficient DNA binding of the soluble transcription factor fragment
 requires dimerization with another bHLH protein.
 -1- SUBCELLULAR LOCATION: Integral membrane protein that moves from
 the endoplasmic reticulum to the Golgi in the absence of sterols.
 -1- ALTERNATIVE PRODUCTS:
 Event-Alternative splicing; Named isoforms=3;
 Comment=Additional isoforms seem to exist;
 Name=SREBP-1A;
 IsoId=P36956-1; Sequence=Displayed;
 Name=SREBP-1B;
 IsoId=P36956-2; Sequence=VSP_002150;
 Name=SREBP-1C;
 IsoId=P36956-3; Sequence=VSP_002149, VSP_002150;
 -1- TISSUE SPECIFICITY: Expressed in a wide variety of tissues, most
 abundant in liver and adrenal gland. In fetal tissues lung and
 liver shows highest expression. Isoform SREBP-1C predominates in
 liver, adrenal gland and ovary, whereas SREBP-1A predominates in
 hepatoma cell lines. Both isoforms are found in kidney, brain,
 white fat, and muscle.
 -1- PTM: Under sterol-depleted conditions, SREBPs are cleaved
 sequentially by site-1 and site-2 protease. The first cleavage
 occurs within the luminal loop and is directly regulated by
 sterols. The second cleavage by site-2 protease occurs within the
 first transmembrane stretch and liberates the soluble
 transcription factor. Cleavage by the cysteine proteases, caspase-
 3 and caspase-7, is induced during apoptosis, independent of
 sterol levels.
 -1- MISCELLANEOUS: SREBPs have to be in a complex with the cleavage-
 activating protein (SCAP) to move to the Golgi and be cleaved by
 site-1 protease.
 -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; U09368; AAC50051.2; -;
 EMBL; S66167; AAB28522.2; ALT_INIT.

DR EMBL; S66168; AAB28523.1; -;
 DR PDB; 1X9; X-ray; A/B/C/D=319-400.
 DR TRANSFAC; T01556; -;
 DR Genew; HGNC:11289; SREBF1.
 DR MIM; 184756; -;
 DR GO; GO:0005789; C:endoplasmic reticulum membrane; TAS.
 DR GO; GO:0005635; C:nuclear membrane; TAS.
 DR GO; GO:0003702; F:RNA polymerase II transcription factor acti. .; TAS.
 DR GO; GO:0006629; P:lipid metabolism; TAS.
 DR GO; GO:0006357; P:regulation of transcription from Pol II pro. .; TAS.
 DR InterPro; IPR001092; HLH_basic.
 DR Pfam; PF00010; HLH; 1.
 DR SMART; SM00353; HLH; 1.
 DR PROSITE; PS50888; HLH; 1.
 KW 3D-structure; Activator; Alternative splicing; Cholesterol metabolism;
 KW Direct protein sequencing; DNA-binding; Endoplasmic reticulum;
 KW Golgi stack; Lipid metabolism; Nuclear protein;
 KW Transcription regulation; Transmembrane.
 FT DOMAIN 1 487 Cytoplasmic (Potential).
 FT TRANSMEM 488 508 Potential.
 FT DOMAIN 509 547 Luminal (Potential).
 FT TRANSMEM 548 568 Potential.
 FT DOMAIN 569 1147 Cytoplasmic (Potential).
 FT DOMAIN 1 60 TRANSCRIPTIONAL ACTIVATION (ACIDIC).
 FT DOMAIN 61 178 Pro/Ser-rich.
 FT DNA_BIND 323 336 Basic motif.
 FT DOMAIN 337 374 Helix-loop-helix motif (By similarity).
 FT DOMAIN 373 394 Leucine-zipper (Potential).
 FT DOMAIN 427 462 Gly/Pro/Ser-rich.
 FT SITE 460 461 Cleavage (by caspase-3 and caspase-7) (By
 similarity).
 FT SITE 490 491 Cleavage (by S2P) (By similarity).
 FT SITE 530 531 Cleavage (by S1P) (Probable).
 FT VARSPLIC 1 29 MDEPPFSAALQALGPPCDLDAALLTDI -> MDCTF
 (in isoform SREBP-1C).
 FT VARSPLIC 1035 1147 /FTid=VSP_002149.
 VFLHEATARINAGASPTRTHQLLDRLSRRRAGPGKGGAVA
 ELEPRTRREHAELALASCYLPFFLSAPQGRVGLAAEA
 RTLEKIGDRLLHDCQQLMRLGGTGVTTSS -> LMDVLT
 SESAWALPFLHKGKGFSPSGHKVPGWHGRMD (in
 isoform SREBP-1B and isoform SREBP-1C).
 /FTid=VSP_002150.
 S->A: No effect on the cleavage by the
 sterol-regulated protease.
 D->A: No effect on the cleavage by the
 sterol-regulated protease.
 S->A: No effect on the cleavage by the
 sterol-regulated protease.
 D->A: No effect on the cleavage by the
 sterol-regulated protease.
 D->A: No effect on the cleavage by the
 sterol-regulated protease.
 G->A: No effect on the cleavage by the
 sterol-regulated protease.
 M->A: No effect on the cleavage by the
 sterol-regulated protease.
 L->A: No effect on the cleavage by the
 sterol-regulated protease.
 D->A: Abolishes the cleavage by the
 sterol-regulated protease.
 R->A: No effect on the cleavage by the
 sterol-regulated protease.
 DRSR->AS: Reduces severely the cleavage
 by the sterol-regulated protease.
 R->A: Abolishes the cleavage by the
 sterol-regulated protease.

Query Match 100.0%; Score 140; DB 1; Length 1147;
 Best Local Similarity 100.0%; Pred. No. 1e-10;

OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
NC	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Uterus;
RC	MEDLINE=22388257; PubMed=12477932;
RX	Klausberg R.B., Feingold E.A., Grouse L.H., Derge J.G.,
RA	Strausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA	Klausner R.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA	Altschul S.F., Jordan H., Moore T., Max S.I., Wang J., Haieh F.,
RA	Hopkins R.F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA	Diatchenko L., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA	Raha S.S., Loquellano M.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA	Phery J., Helton E., Kettner M., Madan A.C., Rodrigues S., Sanchez A.,
RA	Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA	Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA	Jones S.J., Marra M.A.;
RT	"Generation and initial analysis of more than 15,000 full-length human
RT	and mouse cDNA sequences.";
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RL	[2]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Uterus;
RC	Strausberg R.;
RRL	Submitted (SPP-2003) to the EMBL/GenBank/DBJ databases.
DR	EMBL: BC057388; AAS57388.1; -;
DSQ	SEQUENCE 1147 AA; 121674 MW; 58F28870739FF259 CRC64;
Query Match 100.0%; Score 140; DB 2; Length 1147;	
Best Local Similarity 100.0%; Pred. No. 1e-10;	
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps	
QY	1 ARGDPAAQQLWALRALGRPLPTSH 27
DB	594 ARGDPAAQQLWALRALGRPLPTSH 620
RESULT 8	
Q6RIB9	PRELIMINARY; PRT; 1151 AA.
ID	Q6RIB9;
AC	Q6RIB9;
DT	05-JUL-2004 (TrEMBLrel. 27, Created)
DT	05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT	05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
GN	Adipocyte determination and differentiation-dependent factor 1.
DE	Name=ADD1;
OS	Sus scrofa (pig).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX	NCBI_TaxID=9923;
NC	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=Liver;
RC	Li C., Meng H., Zhao W., Pan Y.;
RL	Submitted (DSC-2003) to the EMBL/GenBank/DBJ databases.
RA	EMBL: AY496867; AAS18238.1; -;
DR	InterPro; IPR001092; HLH_basic.
DR	InterPro; IPR000566; Lipoclin_cytfabb.
DR	Pfam; PF00010; HLH; 1.
DR	SMART; SM00353; HLH; 1.
DR	PROSITE; PS00888; HLH; 1.
DR	PROSITE; PS00213; LIPOCALIN; UNKNOWN 1.
DSQ	SEQUENCE 1151 AA; 121501 MW; F9AE7FCAF831F2D CRC64;
Query Match 100.0%; Score 140; DB 2; Length 1151;	

```

RT and mouse cDNA sequences." ;
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RR
RC TISSUE=Brain;
RL SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
RR EMBL; BC063281; AAHG3281.1; -.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00888; HLH; 1.
SQ SEQUENCE 1177 AA; 124633 MW; 4E78AA2358B8C320 CRC64;

Query Match 100.0%; Score 140; DB 2; Length 1177;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 ARGDFAAQAQQLWLALRALGRPLPTSH 27
DB 624 ARGDFAAQAQQLWLALRALGRPLPTSH 650
|||||
|||||

RESULT 11
AAHG3281
ID AAHG3281 PRELIMINARY; PRT; 1177 AA.
AC AAHG3281;
DT 02-WAR-2004 (TReMBLrel. 27, Created)
DT 02-WAR-2004 (TReMBLrel. 27, Last sequence update)
DT 02-WAR-2004 (TReMBLrel. 27, Last annotation update)
DE SREBF1 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diatchenko L., Marasina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettner M.M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grigwold J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences." ;
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RR
RC SEQUENCE FROM N.A.
RL TISSUE=Brain;
RA Strausberg R.;
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
RR EMBL; BC063281; AAHG3281.1; -.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00888; HLH; 1.
SQ SEQUENCE 1177 AA; 124633 MW; 4E78AA2358B8C320 CRC64;

Query Match 100.0%; Score 140; DB 2; Length 1177;
Best Local Similarity 100.0%; Pred. No. 1.1e-10;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1 ARGDFAAQAQQLWLALRALGRPLPTSH 27
|||||
|||||

```

Db 624 ARGDFAAQAQQLWLALRALGRPLPTSH 650

RESULT 12

SREI_RAT STANDARD; PRT; 1024 AA.

AC P56720; Q99PI6; Q99PI7;

DT 30-MAY-2000 (Rel. 39, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Sterol regulatory element binding protein-1 (SREBP-1) (Sterol regulatory element binding transcription factor 1) (Adipocyte determination and differentiation-dependent factor 1) (ADD1) (Fragment).

DE Name=SREBP1; Synonyms=SREBP1;

GN Rattus norvegicus [Rat].

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. [1]

OX NCB1_TaxID=10116;

RN SEQUENCE FROM N.A.

RP TISSUE=Adipocyte;

RC MEDLINE=93330269; PubMed=8336713;

RA Tontonoz P., Kim J.B., Graves R.A., Spiegelman B.M.;

RA "ADD1: a novel helix-loop-helix transcription factor associated with adipocyte determination and differentiation.";

RL Mol. Cell. Biol. 13:4753-4759(1993).

[2]

RN SEQUENCE OF 1-1013 FROM N.A., AND VARIANT MET-755.

RC STRAIN=BN-Lx/Cub, and SHR/Ola;

RC MEDLINE=21212663; PubMed=11309661; DOI=10.1007/s003350010273;

RA Pravenec M., Jansa P., Kostka V., Zidek V., Kren V., Forejt J., Kutz T.W.;

RA "Identification of a mutation in ADD1/SREBP-1 in the spontaneously hypertensive rat.";

RL Mamm. Genome 12:295-298(2001).

[3]

RN DNA-BINDING.

RP MEDLINE=95257939; PubMed=7739539;

RA Kim J.B., Spotts G.D., Halvorsen Y.D., Shih H.M., Ellenberger T., Towie H.C., Spiegelman B.M.;

RA "Dual DNA binding specificity of ADP1/SREBP1 controlled by a single amino acid in the basic helix-loop-helix domain.";

RL Mol. Cell. Biol. 15:2582-2588(1995).

CC -!- FUNCTION: Transcriptional activator that binds to the sterol regulatory element 1 (SRE-1) (5'-ATCACCCAC-3'). Has dual sequence specificity, binding to both an E-box motif (5'-ATCACGGA-3') and to SRE-1 (5'-ATCACCCAC-3'). Regulates the transcription of genes for sterol biosynthesis and the LDL receptor gene.

CC -!- SUBUNIT: Forms a tight complex with SCAP in the ER membrane. Efficient DNA binding of the soluble transcription factor fragment requires dimerization with another bHLH protein.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein that moves from the endoplasmic reticulum to the Golgi in the absence of sterols.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;

CC Name=SREBP-1C; Synonyms=ADD1;

CC IsoId=P56720-1; Sequence=Displayed;

CC Name=SREBP-1A;

CC IsoId=P56720-2; Sequence=Not described;

CC -!- TISSUE SPECIFICITY: Expressed predominantly in brown adipose tissue.

CC -!- PFM: Under sterol-depleted conditions, SREBPs are cleaved sequentially by site-1 and site-2 protease. The first cleavage occurs within the luminal loop and is directly regulated by sterols. The second cleavage by site-2 protease occurs within the first transmembrane stretch and liberates the soluble transcription factor. Cleavage by the cysteine proteases, caspase-3 and caspase-7, is induced during apoptosis, independent of sterol levels (By similarity).

CC -!- MISCELLANEOUS: SREBPs have to be in a complex with the cleavage-activating protein (SCAP) to move to the Golgi and be cleaved by site-1 protease.

CC

CC -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.

CC -!- CAUTION: Ref.1 sequence is a conceptual translation, many probable frame-shifts were corrected from position 878 onward.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).

CC

CC EMBL; L16995; -, NOT ANNOTATED CDS.

DR EMBL; AF286470; AAG28734.2; AUT_TERM.

DR EMBL; AF286469; AAG28733.2; -.

DR HSP; P36956; IAW9.

DR RGD; 69423; Sreb1.

DR InterPro; IPR001092; HLH_basic.

DR Pfam; PF00010; HLH; 1.

DR SMART; SM00353; HLH; 1.

DR PROSITE; PSS0888; HLH; 1.

KW Activator; Alternative splicing; Cholesterol metabolism; DNA-binding; Endoplasmic reticulum; Golgi stack; Lipid metabolism; Nuclear protein; Polymorphism; Transcription regulation; Transmembrane.

FT DOMAIN 1 453

FT TRANSMEM 454 474

FT DOMAIN 475 512

FT TRANSMEM 513 533

FT DOMAIN 534 1023

FT DNA BIND 293 306

FT DOMAIN 307 344

FT DOMAIN 343 364

FT DOMAIN 396 429

FT SITE 427 428

FT SITE 456 457

FT SITE 495 496

FT VARIANT 755 755

FT CONFLICT 243 243

FT CONFLICT 451 451

FT CONFLICT 793 793

FT CONFLICT 950 950

FT NON TER 1024 1024

SQ SEQUENCE 1024 AA; 108826 MW; 29A2D04E5A27404B CRC64;

Query Match 92.1%; Score 129; DB 1; Length 1024;

Best Local Similarity 92.6%; Pred. No. 3.2e-09;

Matches 25; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWLALRALGRPLPTSH 27

Db 559 ARGDFAAQAQQLWLALRALGRPLPTSN 585

RESULT 13

SREI_CRIGR

ID SREI_CRIGR STANDARD; PRT; 1133 AA.

AC Q60416;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE Sterol regulatory element binding protein-1 (SREBP-1) (Sterol regulatory element binding transcription factor 1).

GN Name=SREBP1; Synonyms=SREBP1;

OC Cricetus griseus (Chinese hamster).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.

OX NCB1_TaxID=10029;

[1]

RP SEQUENCE FROM N.A.

```

RX MEDLINE=94274723; PubMed=8006035;
RA Sato R., Yang J., Wang X., Evans M.J., Ho Y.K., Goldstein J.L.,
RT "Assignment of the membrane attachment, DNA binding, and
RT transcriptional activation domains of sterol regulatory element-
RL J. Biol. Chem. 269:17267-17273(1994).
CC -1- FUNCTION: Transcriptional activator that binds to the sterol
CC regulatory element 1 (SRE-1) (5'-ATCACCCAC-3'). Has dual sequence
CC specificity, binding to both an E-box motif (5'-ATCACGTGA-3') and
CC to SRE-1 (5'-ATCACCCAC-3'). Regulates the transcription of genes
CC for sterol biosynthesis and the LDL receptor gene.
CC -1- SUBUNIT: Forms a tight complex with SCAP in the ER membrane.
CC Efficient DNA binding of the soluble transcription factor fragment
CC requires dimerization with another bHLH protein.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein that moves from
CC the endoplasmic reticulum to the Golgi in the absence of sterols.
CC -1- PTM: Under sterol-depleted conditions, SREBPs are cleaved
CC sequentially by site-1 and site-2 protease. The first cleavage
CC occurs within the luminal loop and is directly regulated by
CC sterols. The second cleavage by site-2 protease occurs within the
CC first transmembrane stretch and liberates the soluble
CC transcription factor. Cleavage by the cysteine proteases, caspase-
CC 3 and caspase-7, is induced during apoptosis, independent of
CC sterol levels.
CC -1- MISCELLANEOUS: SREBPs have to be in a complex with the cleavage-
CC activating protein (SCAP) to move to the Golgi and be cleaved by
CC site-1 protease.
CC -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
DR EMBL; U09103; AAA20085.1; -.
DR PIR; A54164; A54164.
DR HSSP; P36956; IAM9.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00888; HLH; 1.
KW Activator; Cholesterol metabolism; DNA-binding; Endoplasmic reticulum;
KW Golgi stack; Lipid metabolism; Nuclear protein;
KW Transcription regulation; Transmembrane.
FT DOMAIN 1 476
FT TRANSMEM 477 497
FT DOMAIN 498 535
FT TRANSMEM 536 556
FT DOMAIN 557 1133
FT DOMAIN 1 60
FT DOMAIN 61 176
FT DNA_BIND 317 330
FT DOMAIN 331 368
FT DOMAIN 367 388
FT DOMAIN 421 454
FT SITE 452 453
FT SITE 479 480
FT SITE 518 519
FT SITE 519 520
SQ SEQUENCE 1133 AA; 120465 MW; 0A77E09DEDCDA84 CRC64;

Query Match 92.1%; Score 129; DB 1; Length 1133;
Best Local Similarity 92.6%; Pred. No. 3.6e-09;
Matches 25; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWALRALGRPLPTSH 27
DB 582 ARGDFAAQAQQLWALRALGRPLPTSN 608

```

```

RESULT 14
Q99JK7 PRELIMINARY; PRT; 866 AA.
ID Q99JK7
AC Q99JK7 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Srebf1 protein (Fragment).
GN Name=Srebf1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Moore T., Max S.I., Wang J., Heide F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
RA Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH II;
RX TISSUE=Mammary tumor metastatized to lung. Tumor arose spontaneously;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006051; AAH06051.1; -.
DR MGI; MGI:107606; Srebf1.
DR GO; GO:0005634; C:nucleus; IDA.
DR GO; GO:0003700; P:transcription factor activity; IDA.
DR GO; GO:0045449; P:regulation of transcription; IDA.
DR InterPro; IPR001092; HLH_basic.
DR Pfam; PF00010; HLH; 1.
DR SMART; SM00353; HLH; 1.
DR PROSITE; PS00888; HLH; 1.
FT NON_TER 1
FT SEQUENCE 866 AA; 93425 MW; CBF6983749B1FC15 CRC64;

Query Match 88.6%; Score 124; DB 2; Length 866;
Best Local Similarity 88.9%; Pred. No. 1.4e-08;
Matches 24; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ARGDFAAQAQQLWALRALGRPLPTSH 27
DB 315 ARGDFAAQAQQLWALRALGRPLPTSN 341

RESULT 15
SREI_MOUSE STANDARD; PRT; 1134 AA.
ID SREI_MOUSE
AC Q9WTN3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)

```


Job time : 31.5417 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:41:17 ; Search time 8.9318 Seconds
(without alignments)
222.664 Million cell updates/sec

Title: US-10-092-750-23

Perfect score: 164

Sequence: 1 GSSKDLAKHIQVVCDDMLTPKIHDLKPQC 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	152	92.7	928	2	US-08-841-483-4
2	152	92.7	928	3	US-03-382-911-4
3	152	92.7	1117	2	US-08-841-483-6
4	152	92.7	1117	3	US-03-382-911-6
5	118	72.0	1065	3	US-09-412-545-2
6	54	32.9	409	4	US-09-540-236-2870
7	51	31.1	387	4	US-09-248-796A-15653
8	50	30.5	46	3	US-08-936-165A-527
9	50	30.5	716	4	US-03-868-758-2
10	50	30.5	716	4	US-03-868-758-41
11	50	30.5	716	4	US-09-868-758-42
12	50	30.5	716	4	US-09-868-758-43
13	50	30.5	716	4	US-09-868-758-44
14	50	30.5	716	4	US-09-868-758-45
15	49	29.9	377	4	US-09-935-038A-2
16	49	29.9	377	4	US-10-094-474-2
17	49	29.9	378	4	US-09-935-038A-5
18	47	28.7	159	4	US-09-270-767-33792
19	47	28.7	159	4	US-09-270-767-49009
20	47	28.7	197	4	US-08-752-165-97
21	47	28.7	565	4	US-03-328-352-4920
22	46	28.0	554	3	US-09-564-805-236
23	46	28.0	573	4	US-09-540-236-2187
24	46	28.0	1006	4	US-08-023-905A-12
25	45.5	27.7	145	2	US-08-467-603-61
26	45.5	27.7	145	2	US-08-466-793-61
27	45.5	27.7	145	2	US-08-491-861A-61

Sequence 61, Appli
Sequence 4, Appli
Sequence 441, App
Sequence 4406, Ap
Sequence 47, Appl
Sequence 47, Appl
Sequence 47, Appl
Sequence 47, Appl
Sequence 5260, Ap
Sequence 23, Appl
Sequence 5267, Ap
Sequence 7404, Ap
Sequence 20231, A
Sequence 2566, Ap
Sequence 84, Appl
Sequence 86, Appl
Sequence 88, Appl

28 45.5 27.7 145 4 US-09-374-671A-61
29 45.5 27.7 341 4 US-03-800-170-4
30 45 27.4 184 4 US-03-198-452A-441
31 45 27.4 424 4 US-09-543-681A-4406
32 44.5 27.1 117 2 US-08-557-309B-47
33 44.5 27.1 117 3 US-08-834-306-47
34 44.5 27.1 117 3 US-08-993-674A-47
35 44.5 27.1 117 4 US-09-256-976-47
36 44.5 27.1 148 4 US-03-583-110-5260
37 44.5 27.1 317 4 US-03-684-708A-23
38 44.5 27.1 326 4 US-03-107-532A-5267
39 44.5 27.1 411 4 US-09-543-681A-7404
40 44 26.8 377 4 US-09-248-796A-20231
41 44 26.8 543 4 US-09-540-236-2566
42 44 26.8 4654 3 US-08-476-515A-84
43 44 26.8 4655 3 US-08-652-877-84
44 44 26.8 4655 3 US-08-652-877-86
45 44 26.8 4655 3 US-08-652-877-88

ALIGNMENTS

RESULT 1

US-08-841-483-4
; Sequence 4, Application US/08841483B
; Patent No. 5976875
; GENERAL INFORMATION:
; APPLICANT: Prescott, Steven M.
; APPLICANT: Bunting, Michaeline
; APPLICANT: Tang, Wen
; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
; TITLE OF INVENTION: Methods of Use Thereof
; FILE REFERENCE: 2037.2.1a
; CURRENT APPLICATION NUMBER: US/08/841,483B
; CURRENT FILING DATE: 1997-04-22
; EARLIER APPLICATION NUMBER: 60/016,210
; EARLIER FILING DATE: 1996-04-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-841-483-4

Query Match 92.7%; Score 152; DB 2; Length 928;
Best Local Similarity 93.3%; Pred. No. 1.6e-14;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSSKDLAKHIQVVCDDMLTPKIHDLKPQC 30
Db 495 GSSKDLAKHIVVCDDMLTPKIQDLKPQC 524

RESULT 2

US-09-382-911-4
; Sequence 4, Application US/09382911
; Patent No. 6221658
; GENERAL INFORMATION:
; APPLICANT: Prescott, Steven M.
; APPLICANT: Bunting, Michaeline
; APPLICANT: Tang, Wen
; APPLICANT: Topnam, Matthew
; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
; TITLE OF INVENTION: Methods of Use Thereof
; FILE REFERENCE: 2037.2.1a
; CURRENT APPLICATION NUMBER: US/09/382,911
; CURRENT FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 08/841,483
; PRIOR FILING DATE: 1997-04-22
; PRIOR APPLICATION NUMBER: 60/016,210

```
; PRIOR FILING DATE: 1996-04-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 4
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-382-911-4

Query Match          92.7%; Score 152; DB 3; Length 928;
Best Local Similarity 93.3%; Pred. No. 1.6e-14;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSSKDLAKHIQVVCDDMDLTPKIHDLKPQC 30
Db 495 GSSKDLAKHIRVVCDDMDLTPKIQDLKPQC 524

RESULT 3
US-08-841-483-6
; Sequence 6, Application US/08841483B
; Patent No. 5976875
; GENERAL INFORMATION:
; APPLICANT: Prescott, Steven M.
; APPLICANT: Bunting, Michaeline
; APPLICANT: Tang, Wen
; APPLICANT: Topham, Matthew
; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
; FILE REFERENCE: 2037.2.1a
; CURRENT APPLICATION NUMBER: US/08/841.483B
; EARLIER FILING DATE: 1997-04-22
; EARLIER FILING DATE: 1996-04-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1117
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-841-483-6

Query Match          92.7%; Score 152; DB 2; Length 1117;
Best Local Similarity 93.3%; Pred. No. 2e-14;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSSKDLAKHIQVVCDDMDLTPKIHDLKPQC 30
Db 684 GSSKDLAKHIRVVCDDMDLTPKIQDLKPQC 713

RESULT 4
US-09-382-911-6
; Sequence 6, Application US/09382911
; Patent No. 6221658
; GENERAL INFORMATION:
; APPLICANT: Prescott, Steven M.
; APPLICANT: Bunting, Michaeline
; APPLICANT: Tang, Wen
; APPLICANT: Topham, Matthew
; TITLE OF INVENTION: Diacylglycerol Kinase Isoforms Epsilon and Zeta and
; FILE REFERENCE: 2037.2.1a
; CURRENT APPLICATION NUMBER: US/09/382.911
; PRIOR FILING DATE: 1999-08-25
; PRIOR FILING DATE: 1997-04-22
; PRIOR FILING DATE: 1996-04-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1117
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-382-911-6

Query Match          92.7%; Score 152; DB 3; Length 1117;
Best Local Similarity 93.3%; Pred. No. 2e-14;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSSKDLAKHIQVVCDDMDLTPKIHDLKPQC 30
Db 684 GSSKDLAKHIRVVCDDMDLTPKIQDLKPQC 713

RESULT 5
US-09-412-545-2
; Sequence 2, Application US/09412545
; Patent No. 6255095
; GENERAL INFORMATION:
; APPLICANT: Prescott, Stephen M.
; APPLICANT: Ding, Li
; APPLICANT: Traer, Elie
; TITLE OF INVENTION: HUMAN DIACYLGLYCEROL KINASE IOTA
; FILE REFERENCE: 1321.2.25
; CURRENT APPLICATION NUMBER: US/09/412.545
; EARLIER FILING DATE: 1999-10-05
; EARLIER FILING DATE: 1998-10-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1065
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-412-545-2

Query Match          72.0%; Score 118; DB 3; Length 1065;
Best Local Similarity 72.4%; Pred. No. 2.9e-09;
Matches 21; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SSKDLAKHIQVVCDDMDLTPKIHDLKPQC 30
Db 576 SSRDLKSHVKKVVCDDTDLTPKIQELKPQC 604

RESULT 6
US-09-540-236-2870
; Sequence 2870, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRI
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540.236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2870
; LENGTH: 409
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2870

Query Match          32.9%; Score 54; DB 4; Length 409;
Best Local Similarity 40.7%; Pred. No. 5;
Matches 11; Conservative 7; Mismatches 7; Indels 2; Gaps 1;

Qy 4 KDLA--KHIVVCDDMDLTPKIHDLKP 28
Db 9 KDLGMDKHITQLICDPQVPMGINSIQP 35

RESULT 7
US-09-248-796A-15653
```

```

; TELEX:
; INFORMATION FOR SEQ ID NO: 527:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-936-165A-527

Query Match 30.5%; Score 50; DB 3; Length 46;
Best Local Similarity 32.1%; Pred. No. 1.5;
Matches 9; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 2 SSKDLAKHIQVVDGMDLTPKIHDLKPQ 29
DB 13 ANKVDKQVQALIDEIDRNPNTDKEKQ 40

RESULT 9
US-09-868-758-2
; Sequence 2, Application US/09868758
; Patent No. 6576439
; GENERAL INFORMATION:
; APPLICANT: Glaxo Wellcome KK
; APPLICANT: Takemoto, Yoshihiro
; APPLICANT: Sakai, Yutaka
; APPLICANT: Hashimoto, Yasuhiro
; TITLE OF INVENTION: IKK3
; FILE REFERENCE: 9950986P
; CURRENT APPLICATION NUMBER: US/09/868,758
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: GB 9828704.8
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-868-758-2

Query Match 30.5%; Score 50; DB 4; Length 716;
Best Local Similarity 56.2%; Pred. No. 40;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 KDLAKHIQVVDGMDL 19
DB 677 KDLLEHMQELCEGMKL 692

RESULT 10
US-09-868-758-41
; Sequence 41, Application US/09868758
; Patent No. 6576439
; GENERAL INFORMATION:
; APPLICANT: Glaxo Wellcome KK
; APPLICANT: Takemoto, Yoshihiro
; APPLICANT: Sakai, Yutaka
; APPLICANT: Hashimoto, Yasuhiro
; TITLE OF INVENTION: IKK3
; FILE REFERENCE: 9950986P
; CURRENT APPLICATION NUMBER: US/09/868,758
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: GB 9828704.8
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 41
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:

```

; OTHER INFORMATION: Description of Artificial Sequence: DT7-1KK3
; OTHER INFORMATION: mutant
US-09-868-758-41

Query Match 30.5%; Score 50; DB 4; Length 716;
Best Local Similarity 56.2%; Pred. No. 40;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 KDLAKHIQVVCDDMDL 19
||| |:::|
DB 677 KDLHMQELCEGMKL 692

RESULT 11
US-09-868-758-42
; Sequence 42, Application US/09868758
; Patent No. 6576439
; GENERAL INFORMATION:
; APPLICANT: Glaxo Wellcome KK
; APPLICANT: Takemoto, Yoshihiro
; APPLICANT: Sakai, Yutaka
; APPLICANT: Hashimoto, Yasuhiro
; TITLE OF INVENTION: IKK3
; FILE REFERENCE: 9950986P
; CURRENT APPLICATION NUMBER: US/09/868,758
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: GB 9828704.8
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 42
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DT7-1KK3
; OTHER INFORMATION: mutant
US-09-868-758-42

Query Match 30.5%; Score 50; DB 4; Length 716;
Best Local Similarity 56.2%; Pred. No. 40;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 KDLAKHIQVVCDDMDL 19
||| |:::|
DB 677 KDLHMQELCEGMKL 692

RESULT 12
US-09-868-758-43
; Sequence 43, Application US/09868758
; Patent No. 6576439
; GENERAL INFORMATION:
; APPLICANT: Glaxo Wellcome KK
; APPLICANT: Takemoto, Yoshihiro
; APPLICANT: Sakai, Yutaka
; APPLICANT: Hashimoto, Yasuhiro
; TITLE OF INVENTION: IKK3
; FILE REFERENCE: 9950986P
; CURRENT APPLICATION NUMBER: US/09/868,758
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: GB 9828704.8
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DT7-1KK3
; OTHER INFORMATION: mutant
US-09-868-758-43

Query Match 30.5%; Score 50; DB 4; Length 716;
Best Local Similarity 56.2%; Pred. No. 40;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 KDLAKHIQVVCDDMDL 19
||| |:::|
DB 677 KDLHMQELCEGMKL 692

RESULT 13
US-09-868-758-44
; Sequence 44, Application US/09868758
; Patent No. 6576439
; GENERAL INFORMATION:
; APPLICANT: Glaxo Wellcome KK
; APPLICANT: Takemoto, Yoshihiro
; APPLICANT: Sakai, Yutaka
; APPLICANT: Hashimoto, Yasuhiro
; TITLE OF INVENTION: IKK3
; FILE REFERENCE: 9950986P
; CURRENT APPLICATION NUMBER: US/09/868,758
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: GB 9828704.8
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DT7-1KK3
; OTHER INFORMATION: mutant
US-09-868-758-44

Query Match 30.5%; Score 50; DB 4; Length 716;
Best Local Similarity 56.2%; Pred. No. 40;
Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 4 KDLAKHIQVVCDDMDL 19
||| |:::|
DB 677 KDLHMQELCEGMKL 692

RESULT 14
US-09-868-758-45
; Sequence 45, Application US/09868758
; Patent No. 6576439
; GENERAL INFORMATION:
; APPLICANT: Glaxo Wellcome KK
; APPLICANT: Takemoto, Yoshihiro
; APPLICANT: Sakai, Yutaka
; APPLICANT: Hashimoto, Yasuhiro
; TITLE OF INVENTION: IKK3
; FILE REFERENCE: 9950986P
; CURRENT APPLICATION NUMBER: US/09/868,758
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: GB 9828704.8
; PRIOR FILING DATE: 1998-12-24
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 45
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DT7-1KK3
; OTHER INFORMATION: mutant
US-09-868-758-45

Query Match 30.5%; Score 50; DB 4; Length 716;
Best Local Similarity 56.2%; Pred. No. 40;

This Page Blank (usp10)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 15:53:52 ; Search time 28,9352 Seconds
(without alignments)
366.225 Million cell updates/sec

Title: US-10-092-750-23

Perfect score: 164

Sequence: 1 GSSKDLAKHIQVCDGMDLTPKIHDLKPOC 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	164	100.0	30	14	US-10-092-750-23
2	152	92.7	584	9	US-09-764-868-1054
3	152	92.7	928	14	US-10-170-385-353
4	152	92.7	929	14	US-10-205-194-7
5	118	72.0	1065	9	US-09-771-161A-239
6	118	72.0	1065	16	US-10-408-765A-582
7	57.5	35.1	60	17	US-10-425-115-294496
8	54	32.9	318	16	US-10-437-963-156582
9	50.5	30.8	732	15	US-10-229-345-3
10	50.5	30.8	732	15	US-10-274-177-3
11	50.5	30.8	732	16	US-10-650-112-3
12	50.5	30.8	755	14	US-10-295-027-124
13	50.5	30.8	755	14	US-10-295-027-1186

14	50.5	30.8	755	15	US-10-173-999-87	Sequence 87, Appl
15	50.5	30.8	755	15	US-10-229-345-14	Sequence 14, Appl
16	50.5	30.8	755	15	US-10-274-177-14	Sequence 14, Appl
17	50.5	30.8	755	16	US-10-650-112-14	Sequence 14, Appl
18	50.5	30.8	807	10	US-09-830-020A-2	Sequence 2, Appli
19	50.5	30.8	807	14	US-10-295-027-122	Sequence 122, App
20	50.5	30.8	807	14	US-10-295-027-764	Sequence 764, App
21	50.5	30.8	807	14	US-10-295-027-777	Sequence 777, App
22	50.5	30.8	807	14	US-10-295-027-883	Sequence 883, App
23	50.5	30.8	807	14	US-10-295-027-1187	Sequence 1187, App
24	50.5	30.8	807	14	US-10-087-080-23	Sequence 23, Appl
25	50.5	30.8	807	15	US-10-173-999-85	Sequence 85, Appl
26	50.5	30.8	826	14	US-10-295-027-120	Sequence 120, Appl
27	50.5	30.8	826	15	US-10-173-999-83	Sequence 83, Appl
28	50	30.5	46	9	US-09-939-980-527	Sequence 527, Appl
29	50	30.5	310	14	US-10-279-029-103	Sequence 103, Appl
30	50	30.5	310	14	US-10-219-810-30	Sequence 30, Appl
31	50	30.5	311	14	US-10-219-810-37	Sequence 37, Appl
32	50	30.5	360	16	US-10-437-963-166581	Sequence 166581,
33	50	30.5	541	14	US-10-369-493-622	Sequence 622, App
34	50	30.5	716	14	US-10-298-402-2	Sequence 2, Appli
35	50	30.5	716	14	US-10-408-636-2	Sequence 2, Appli
36	50	30.5	716	14	US-10-408-636-41	Sequence 41, Appl
37	50	30.5	716	14	US-10-408-636-42	Sequence 42, Appl
38	50	30.5	716	14	US-10-408-636-43	Sequence 43, Appl
39	50	30.5	716	14	US-10-408-636-44	Sequence 44, Appl
40	50	30.5	716	14	US-10-408-636-45	Sequence 45, Appl
41	50	30.5	716	14	US-10-394-322A-34	Sequence 34, Appl
42	50	30.5	2437	9	US-09-815-242-5834	Sequence 5834, Ap
43	50	30.5	3533	15	US-10-282-122A-70177	Sequence 70177, A
44	50	30.5	6281	9	US-09-815-242-12996	Sequence 12996, A
45	50	30.5	6713	15	US-10-282-122A-43811	Sequence 43811, A

ALIGNMENTS

RESULT 1
US-10-092-750-23
; Sequence 23, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 30
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-23

Query Match 100.0%; Score 164; DB 14; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.6e-17;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GSSKDLAKHIQVCDGMDLTPKIHDLKPOC 30
Db 1 GSSKDLAKHIQVCDGMDLTPKIHDLKPOC 30

RESULT 2
US-09-764-868-1054
; Sequence 1054, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P7232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1054
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (50)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (311)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (545)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-1054

```

```

Query Match          92.7%; Score 152; DB 9; Length 584;
Best Local Similarity 93.3%; Pred. No. 7e-14;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSSKDLAKHIQVCGMDLTPKIHDLKPQC 30
Db 195 GSSKDLAKHIRVCGMDLTPKIQDLKPQC 224

```

```

RESULT 3
US-10-170-385-353
; Sequence 353, Application US/10170385
; Publication No. US20030203372A1
; GENERAL INFORMATION:
; APPLICANT: Ward, Neil Raymond
; APPLICANT: Mundy, Christopher Robert
; APPLICANT: Kan, On
; APPLICANT: Harris, Robert Alan
; APPLICANT: White, Jonathan
; APPLICANT: Binley, Katie Mary
; APPLICANT: Rayner, William Nigel
; APPLICANT: Naylor, Stuart
; APPLICANT: Kingsman, Susan Mary
; APPLICANT: Krige, David
; TITLE OF INVENTION: ANALYSIS METHOD
; FILE REFERENCE: 532682000100
; CURRENT APPLICATION NUMBER: US/10/170,385
; CURRENT FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: PCT/GS02/01662
; PRIOR FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: PCT/GS01/05458
; PRIOR FILING DATE: 2001-12-10
; NUMBER OF SEQ ID NOS: 549
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 353
; LENGTH: 928
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-170-385-353

Query Match          92.7%; Score 152; DB 14; Length 928;
Best Local Similarity 93.3%; Pred. No. 1.2e-13;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSSKDLAKHIQVCGMDLTPKIHDLKPQC 30
Db 495 GSSKDLAKHIRVCGMDLTPKIQDLKPQC 524

```

```

RESULT 4
US-10-205-194-7
; Sequence 7, Application US/10205194
; Publication No. US20030134301A1
; GENERAL INFORMATION:
; APPLICANT: Warner-Lambert Company
; APPLICANT: Lee, Kevin
; APPLICANT: Dixon, Alistair
; APPLICANT: Brooksbank, Robert
; APPLICANT: Finnock, Robert
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain
; FILE REFERENCE: WL-A-018201
; CURRENT APPLICATION NUMBER: US/10/205,194
; CURRENT FILING DATE: 2000-07-24
; PRIOR APPLICATION NUMBER: GB 0118354.0
; PRIOR FILING DATE: 2001-07-27
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 929
; TYPE: PRT
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Diacylglycerol kinase
US-10-205-194-7

Query Match          92.7%; Score 152; DB 14; Length 929;
Best Local Similarity 93.3%; Pred. No. 1.2e-13;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GSSKDLAKHIQVCGMDLTPKIHDLKPQC 30
Db 496 GSSKDLAKHIRVCGMDLTPKIQDLKPQC 525

```

```

RESULT 5
US-09-771-161A-239
; Sequence 239, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 239
; LENGTH: 1065
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-239

Query Match          72.0%; Score 118; DB 9; Length 1065;
Best Local Similarity 72.4%; Pred. No. 1.7e-08;
Matches 21; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SSKDLAKHIQVCGMDLTPKIHDLKPQC 30
Db 576 SSRDLAKHIRVCGMDLTPKIQDLKPQC 604

RESULT 6
US-10-408-765A-582
; Sequence 582, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:

```


APPLICANT: Ghosh, Soumitra S.
APPLICANT: Fahy, Eoin D.
APPLICANT: Zhang, Bing
APPLICANT: Gibson, Bradford W.
APPLICANT: Taylor, Steven W.
APPLICANT: Glenn, Gary M.
APPLICANT: Warnock, Dale E.
TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
FILE REFERENCE: 660088.465
CURRENT APPLICATION NUMBER: US/10/408,765A
CURRENT FILING DATE: 2003-04-04
NUMBER OF SEQ ID NOS: 3077
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 582
LENGTH: 1065
TYPE: PRT
ORGANISM: Homo sapiens
US-10-408-765A-582

Query Match 72.0%; Score 118; DB 16; Length 1065;
Best Local Similarity 72.4%; Pred. No. 1.7e-08;
Matches 21; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 SSKDLAKHIQVCDGMDLTPKIHDLKPC 30
DB 576 SSRDLSKHVXVCDGDLTPKIQELKPC 604

RESULT 7
US-10-425-115-294496
Sequence 294496, Application US/10425115
Publication No. US20040214272A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53222)B
CURRENT APPLICATION NUMBER: US/10/425,115
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 369326
SEQ ID NO 294496
LENGTH: 60
TYPE: PRT
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: MRT4577_31667C.1.pep
US-10-425-115-294496

Query Match 35.1%; Score 57.5; DB 17; Length 60;
Best Local Similarity 52.4%; Pred. No. 0.7; Mismatches 2; Indels 1; Gaps 1;
Matches 11; Conservative 7; Mismatches 2; Indels 1; Gaps 1;

QY 9 HIQVCDGMDLTPKIHDLKPC 29
DB 11 NLQVNCNGI-LTSKIHLRPR 30

RESULT 8
US-10-437-963-156582
Sequence 156582, Application US/10437963
Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Barbazuk, Brad

APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53221)B
CURRENT APPLICATION NUMBER: US/10/437,963
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 156582
LENGTH: 318
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_56235C.1.pep
US-10-437-963-156582

Query Match 32.9%; Score 54; DB 16; Length 318;
Best Local Similarity 43.5%; Pred. No. 16;
Matches 10; Conservative 7; Mismatches 6; Indels 0; Gaps 0;

QY 2 SSKDLAKHIQVCDGMDLTPKIH 24
DB 90 ASRDIALDLQVRLNGLDTSRPH 112

RESULT 9
US-10-229-345-3
Sequence 3, Application US/10229345
Publication No. US20040038220A1
GENERAL INFORMATION:
APPLICANT: MARKOWITZ, Sanford D.
TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
FILE REFERENCE: CWRU-P01-003
CURRENT APPLICATION NUMBER: US/10/229,345
CURRENT FILING DATE: 2002-08-26
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 732
TYPE: PRT
ORGANISM: Homo sapiens
US-10-229-345-3

Query Match 30.8%; Score 50.5; DB 15; Length 732;
Best Local Similarity 42.9%; Pred. No. 1.4e-02;
Matches 9; Conservative 7; Mismatches 4; Indels 1; Gaps 1;

QY 3 SKDLAKHIQV-VCDGMDLTPK 22
DB 43 SFERSKHFAITVCDGLDISPE 63

RESULT 10
US-10-274-177-3
Sequence 3, Application US/10274177
Publication No. US20040038225A1
GENERAL INFORMATION:
APPLICANT: MARKOWITZ, Sanford D.
TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
FILE REFERENCE: CWRU-P01-003
CURRENT APPLICATION NUMBER: US/10/274,177
CURRENT FILING DATE: 2002-10-18
PRIOR APPLICATION NUMBER: US/10/229,345
PRIOR FILING DATE: 2002-08-26
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 732
TYPE: PRT
ORGANISM: Homo sapiens
US-10-274-177-3

Query Match 30.8%; Score 50.5; DB 15; Length 732;
Best Local Similarity 42.9%; Pred. No. 1.4e-02;

Matches 9; Conservative 7; Mismatches 4; Indels 1; Gaps 1;

QY 3 SKDLAKHIQV-VCDGMDLTPK 22
 | : | | : | | | : | | :
Db 43 SPERSKHFAITVCDGLDISPE 63

RESULT 11

US-10-650-112-3
; Sequence 3, Application US/10650112
; Publication No. US20040110712A1
; GENERAL INFORMATION:
; APPLICANT: VARKOWITZ, Sanford D.
; TITLE OF INVENTION: METHODS FOR TREATING PATIENTS AND IDENTIFYING THERAPEUTICS
; FILE REFERENCE: CWRU-POL-044
; CURRENT APPLICATION NUMBER: US/10/650,112
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: 10/274,177
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: 10/229,245
; PRIOR FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 60/406,296
; PRIOR FILING DATE: 2002-08-27
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 732
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-650-112-3

Query Match 30.8%; Score 50.5; DB 16; Length 732;
Best Local Similarity 42.9%; Pred. No. 1.4e+02;
Matches 9; Conservative 7; Mismatches 4; Indels 1; Gaps 1;

QY 3 SKDLAKHIQV-VCDGMDLTPK 22
 | : | | : | | | : | | :
Db 43 SPERSKHFAITVCDGLDISPE 63

RESULT 12

US-10-295-027-124
; Sequence 124, Application US/10295027
; Publication No. US2003023250A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynnne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Bos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; METHOD OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211

Db 66 SFERSKHFAITVCDGLDISPE 86

Local Matches	Conservative Matches	Mismatches	Indels	Gaps
9	9	7	1	1

Qy	3 SKDLAKHIQV-VCDGMDLTPK 22
Dd	66 SFERSKHFAITVCDGLDISPE 86

```

, TITLE OF INVENTION: Methods of diagnosis of Ovarian Cancer, Compositions
,
, TITLE OF INVENTION: and Methods of Screening for Modulators of Ovarian
,
, TITLE OF INVENTION: Cancer
,
, FILE REFERENCE: 018501-00242005
,
, CURRENT APPLICATION NUMBER: US/10/173,999
,
, CURRENT FILING DATE: 2002-06-17
,
, PRIOR APPLICATION NUMBER: US 60/299,234
,
, PRIOR FILING DATE: 2001-06-18
,
, PRIOR APPLICATION NUMBER: US 60/315,287
,
, PRIOR FILING DATE: 2001-08-27
,
, PRIOR APPLICATION NUMBER: US 60/350,666
,
, PRIOR FILING DATE: 2001-11-13
,
, PRIOR APPLICATION NUMBER: US 60/372,246
,
, PRIOR FILING DATE: 2001-04-12
,
, NUMBER OF SEQ ID NOS: 163
,
, SOFTWARE: PatentIn Ver. 2.1
,
, SEQ. ID NO 87
,
, LENGTH: 755
,
, TYPE: PRT
,
, ORGANISM: Homo sapiens
,
, US-10-173-999-87

```

Matches	9;	Conservative	7;	Mismatches	4;	Indels	1;	Gaps	1;
----------------	----	---------------------	----	-------------------	----	---------------	----	-------------	----

Qy 3 SKDLAKHIQV-VCDGMDLTPK 22
| : : | : : : : :
Db 66 SFERSKHFAITVCDGLDISPE 86

```

, TITLE OF INVENTION: METHODS FOR CATEGORIZING PATIENTS
,
, FILE REFERENCE: CMRU-P01-003
, CURRENT APPLICATION NUMBER: US/10/229,345
, CURRENT FILING DATE: 2002-08-26
, NUMBER OF SEQ ID NOS: 20
, SOFTWARE: PatentIn version 3.1
, SEQ ID NO 14
, LENGTH: 755
, TYPE: PRT
, ORGANISM: Homo sapiens
US-10-229-345-14

```

Matches	9;	Conservative	7;	Mismatches	4;	Indels	1;	Gaps	1;
---------	----	--------------	----	------------	----	--------	----	------	----

QY 3 SKDLAKHIQV-VCDGMDLTPK 22

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:36:51 ; Search time 6.06481 Seconds
(without alignments)
475.942 Million cell updates/sec

Title: US-10-092-750-23

Perfect score: 164

Sequence: 1 GSSKDLAKHIQVCDGMDLTPKIHDLKPQC 30

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_791*

1: Piri:*

2: Piri2:*

3: Piri3:*

4: Piri4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	152	92.7	929	2 JC6124	diacylglycerol kin
2	84	51.2	930	2 T34334	hypothetical prote
3	52	31.7	262	2 D70895	hypothetical prote
4	51.5	31.4	7829	2 T15789	hypothetical prote
5	51	31.1	818	2 T35872	DNA gyrase-like pr
6	51	31.1	1454	2 T13709	diacylglycerol kin
7	50	30.5	525	2 T18569	inositol-3-phospha
8	50	30.5	541	2 C75358	glucose-6-phospha
9	50	30.5	607	2 T45969	mRNA capping enzym
10	50	30.5	822	2 T25866	hypothetical prote
11	50	30.5	6713	2 B89821	hypothetical prote
12	49	29.9	99	2 B86203	hypothetical prote
13	49	29.9	152	1 NRRT	pancreatic ribonuc
14	49	29.9	186	2 F81689	conserved hypotet
15	49	29.9	188	2 F71534	hypothetical prote
16	49	29.9	592	1 C42594	L-amino acid hydrol
17	48	29.3	101	2 I48148	Neutrophil attract
18	48	29.3	147	2 S74986	regulatory compone
19	48	29.3	203	1 YNECHI	phosphoribosyl-AMP
20	48	29.3	368	2 A69774	integrase homolog
21	48	29.3	517	2 S28229	diacylglycerol kin
22	48	29.3	570	2 F71040	glycine-tRNA ligas
23	48	29.3	571	2 C75174	glycyl-tRNA synthet
24	48	29.3	635	2 D84920	hypothetical prote
25	48	29.3	791	2 A46140	diacylglycerol kin
26	48	29.3	796	2 B46140	diacylglycerol kin
27	48	29.3	1186	2 T33754	O/E-1-associated z
28	47.5	29.0	124	1 NRHP	pancreatic ribonuc
29	47.5	29.0	199	2 JC5718	superoxide dismuta

30 47.5 29.0 269 2 AC0973 lipopolysaccharide
31 47.5 29.0 269 2 D41317 probable lipopolys
32 47 28.7 267 2 G69315 translation initia
33 47 28.7 557 2 T44843 glucose-6-phosphat
34 47 28.7 775 2 T38929 changed division r
35 46.5 28.4 249 2 T37669 hypothetrical zinc-
36 46.5 28.4 313 2 B86908 hypothetrical prote
37 46.5 28.4 421 2 T38811 hypothetrical prote
38 46.5 28.4 505 2 F82396 hypothetrical prote
39 46.5 28.4 594 2 S33561 ref(2)P protein -
40 46 28.0 77 2 A47070 ferrocyclatase (EC
41 46 28.0 124 1 NRCM pancreatic ribonuc
42 46 28.0 124 1 NRCMB pancreatic ribonuc
43 46 28.0 124 1 NRCMW pancreatic ribonuc
44 46 28.0 148 2 S50547 hypothetrical prote
45 46 28.0 157 2 A99226 conserved hypothet

ALIGNMENTS

RESULT 1

JC6124
diacylglycerol kinase (EC 2.7.1.107) IV - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 16-Apr-1997 #sequence_revision 09-May-1997 #text_change 09-Jul-2004
C:Accession: JC6124
R:Goto, K.; Kondo, H.
Proc. Natl. Acad. Sci. U.S.A. 93, 11195-11201, 1996
A:Title: A 104-kDa diacylglycerol kinase containing ankyrin-like repeats localizes in the
A:Reference number: JC6124; MUID:97008156; PMID:8855332
A:Contents: Tissue Retina,Brain
A:Accession: JC6124
A:Molecule type: mRNA
A:Residues: 1-929 <GOT>
A:Cross-references: UNIPROT:O08560; DBJ:D78588; NID:G1906781; PIDN:BAAL8942.1; PID:dl01;
C:Comment: This enzyme is involved in intranuclear processes, it plays a role in the acti
cylglycerol to phosphatidic acid.
C:Keywords: ATP binding; brain; phosphotransferase; zinc finger

Query Match 92.7%; Score 152; DB 2; Length 929;
Best Local Similarity 93.3%; Pred. No. 2.3e-13;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSSKDLAKHIQVCDGMDLTPKIHDLKPQC 30
|||||
DB 496 GSSKDLAKHIRVCDGMDLTPKIQDLKPQC 525
|||||

RESULT 2

T34334
hypothetical protein K06A1.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34334
R:Fulton, L.
submitted to the EMBL Data Library, July 1995
A:Description: The sequence of C. elegans cosmid K06A1.
A:Reference number: Z21507
A:Accession: T34334
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-930 <FUI>
A:Cross-references: EMBL:U23449; PIDN:AAC24300.1; GSPDB:GN000020; CESP:K06A1.6
A:Experimental source: strain Bristol N2; clone K06A1
C:Genetics:
A:Gene: CESP:K06A1.6
A:Map position: 2
A:Introns: 173/3; 265/2; 300/3; 484/3; 537/2; 587/1; 684/3; 796/1; 840/1; 897/2

Query Match 51.2%; Score 84; DB 2; Length 930;
Best Local Similarity 51.7%; Pred. No. 0.001;
Matches 15; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

```

C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T35872
R:Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, November 1997
A:Reference number: Z21592
A:Accession: T35872
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-818 <OLI>
A:Cross-references: UNIPROT:O50510; EMBL:AL009204; PIDN:CAA15793.1; GSPDB:GN00070; SCODE
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODEB:SC9B10.03C
C:Superfamily: DNA topoisomerase (ATP-hydrolyzing) chain A; phage T4 DNA topoisomerase (
Query Match 31.1%; Score 51; DB 2; Length 818;
Best Local Similarity 50.0%; Pred. No. 43;
Matches 13; Conservative 3; Mismatches 8; Indels 2; Gaps 1;
QY 4 KDL--AKHIQVVCDCGMDLTPKIHDLK 27
DB 289 KDLVGAKKIQGIADVKDLTDRAHGLR 314
RESULT 6
T13709
diacylglycerol kinase (EC 2.7.1.107) - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 09-Jul-2004
C:Accession: T13709
R:Masai, I.; Okazaki, A.; Hosoya, T.; Hotta, Y.
Proc. Natl. Acad. Sci. U.S.A. 90, 11157-11161, 1993
A:Title: Drosophila retinal degeneration A gene encodes an eye-specific diacylglycerol ki
A:Reference number: Z17702; MUID:94068563; PMID:8248222
A:Accession: T13709
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1454 <MAS>
A:Cross-references: UNIPROT:Q09103; EMBL:D17315; NID:d1007485; PID:d1004654; PIDN:BAA041:
C:Genetics:
A:Gene: rdga
A:Cross-references: FlyBase:FBgn0003217
C:Keywords: phosphotransferase
Query Match 31.1%; Score 51; DB 2; Length 1454;
Best Local Similarity 36.4%; Pred. No. 77;
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
QY 4 KDLAKHIQVVCDCGMDLTPKIHDLK 25
DB 1013 RNLSQVWLECDGQDFGKLRD 1034
RESULT 7
T18569
inositol-3-phosphate synthase (EC 5.5.1.4) - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18569; T20002
R:Ainscough, R.
submitted to the EMBL Data Library, December 1998
A:Reference number: Z18979
A:Accession: T18569
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-525 <WIL>
A:Cross-references: UNIPROT:Q18664; EMBL:AL033535; PIDN:CAA22132.1; CESP:VF13D12L.1
A:Experimental source: clone VF13D12L
R:Gajadaty, S.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z19209
A:Accession: T20002

```

```

QY 2 SSKDLAKHIQVVCDCGMDLTPKIHDLKQC 30
DB 615 SWKDLCEVITLCEGDVDTPIKELKHC 643
RESULT 3
D70895
hypothetical protein Rv1086 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Jul-2004
C:Accession: D70895
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: D70895
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-262 <COL>
A:Cross-references: UNIPROT:O53434; GB:AL021897; GB:AL123456; NID:G3256022; PIDN:CAA1720
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv1086
C:Superfamily: conserved hypothetical protein YBR002c
Query Match 31.7%; Score 52; DB 2; Length 262;
Best Local Similarity 60.0%; Pred. No. 9.6;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 2 SSKDLAKHIQVVCDCG 16
DB 27 SKSDLPRIHVLCDG 41
RESULT 4
T15789
hypothetical protein C41A3.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T15789
R:Bentley, D.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid C41A3.
A:Reference number: Z18404
A:Accession: T15789
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-7829 <BEN>
A:Cross-references: UNIPROT:Q18559; EMBL:U41541; NID:g1109867; PID:g1109869; PIDN:AAA831
C:Genetics:
A:Gene: CESP:C41A3.1
A:Introns: 29/2; 69/3; 141/2; 192/2; 271/3; 321/2; 1000/1; 1098/1; 1127/3; 1168/2; 1282/
2; 7514/3; 7596/3; 7654/3; 7716/3; 7732/3; 7769/2; 7797/3
C:Keywords: carrier protein; phosphatethione; phosphoprotein
F:6547-6616/Domain: acyl carrier protein homology <AC>
F:12832,5271,6580/Binding site: phosphatethione (ser) (covalent) #status predicted
Query Match 31.4%; Score 51.5; DB 2; Length 7829;
Best Local Similarity 38.5%; Pred. No. 3.6e+02;
Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;
QY 2 SSKDLAKHIQVVCDCGMDLTPK-IHDL 26
DB 4730 SAKSLQYVQVLCBFISAKSLHDI 4755
RESULT 5
T35872
DNA gyrase-like protein - Streptomyces coelicolor

```

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-525 <WI2>

A;Cross-references: EMBL:Z69902; PIDN:CAA93771.1; GSPDB:GN00020; CESP:VF13D12L.1

A;Experimental source: clone C47D12

C;Genetics:

A;Gene: CESP:VF13D12L.1

A;Map position: 2

A;Introns: 106/2; 287/1; 411/2

C;Superfamily: myo-inositol-1-phosphate synthase

C;Keywords: intramolecular lyase; isomerase

Query Match 30.5%; Score 50; DB 2; Length 525;

Best Local Similarity 40.7%; Pred. No. 37;

Matches 11; Conservative 4; Mismatches 6; Indels 6; Gaps 1;

QY 9 HIQVCDGMDLTPKIHDP-----LKPK 29

Db 37 HFHRRADGLHVTPEKHDYFKTVLKPR 63

RESULT 8

C75358

Glucose-6-phosphate isomerase - Deinococcus radiodurans (strain R1)

C;Species: Deinococcus radiodurans

C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C;Accession: C75358

R;White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A;Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A;Reference number: A75250; MUID:20036896; PMID:10567266

A;Accession: C75358

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-541 <WHI>

A;Cross-references: UNIPROT:Q9RTI8; GB:AE002017; GB:AE000513; NID:96459527; PIDN:AAF1129

A;Experimental source: strain R1

C;Genetics:

A;Gene: DR1742

A;Map position: 1

C;Superfamily: glucose-6-phosphate isomerase

Query Match 30.5%; Score 50; DB 2; Length 541;

Best Local Similarity 39.3%; Pred. No. 39;

Matches 11; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 2 SSKDLAKHIQVCDGMDLTPKIHDLKPQ 29

Db 170 AQBDLTVRFVSNVDCITLTKETRLDLP 197

RESULT 9

T45969

mRNA capping enzyme-like protein - Arabidopsis thaliana

N;Alternate names: protein F7J8.270

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 09-Jul-2004

C;Accession: T45969

R;Revan, M.; Zimmermann, W.; Gruenewald, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Lem

submitted to the Protein Sequence Database, January 2000

A;Reference number: Z23018

A;Accession: T45969

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-607 <BEV>

A;Cross-references: UNIPROT:Q9LFA7; EMBL:AL137189

A;Experimental source: cultivar Columbia; BAC clone F7J8

C;Genetics:

A;Map position: 5

A;Introns: 33/3; 73/1; 112/3; 146/3; 167/3; 212/3; 281/3; 321/3; 351/3; 398/3; 434/3; 46

A;Note: F7J8.270

Query Match 30.5%; Score 50; DB 2; Length 607;

Best Local Similarity 44.4%; Pred. No. 43;

Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 4 KDLAKHIQVCDGMDLTP 21

Db 139 KGIKHVKIACGRDAVP 156

RESULT 10

T25866

hypotheical protein T07F8.1 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C;Accession: T25866

R;Pauley, A.

submitted to the EMBL Data Library, December 1996

A;Description: The sequence of C. elegans cosmid T07F8.

A;Reference number: Z20102

A;Accession: T25866

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-822 <PAU>

A;Cross-references: UNIPROT:P91440; EMBL:U80954; PIDN:AAB38096.1; GSPDB:GN00020; CESP:TO:

A;Experimental source: strain Bristol N2; clone T07F8

C;Genetics:

A;Gene: CESP:T07F8.1

A;Map position: 2

A;Introns: 578/3

Query Match 30.5%; Score 50; DB 2; Length 822;

Best Local Similarity 39.1%; Pred. No. 59;

Matches 9; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 4 KDLAKHIQVCDGMDLTPKIHDL 26

Db 736 KNLKRNIRILHPSMGVTPKHL 758

RESULT 11

B89921

hypotheical protein ebhA [imported] - Staphylococcus aureus (strain N315)

C;Species: Staphylococcus aureus

C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004

C;Accession: B89921

R;Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A;Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A;Reference number: A89758; MUID:21311952; PMID:11418146

A;Accession: B89921

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-6713 <KUR>

A;Cross-references: UNIPROT:Q99U54; GB:BA000018; PID:gl3701232; PIDN:BA842527.1; GSPDB:G

A;Experimental source: strain N315

C;Genetics:

A;Gene: ebhA

Query Match 30.5%; Score 50; DB 2; Length 6713;

Best Local Similarity 32.1%; Pred. No. 5.1e+02;

Matches 9; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 2 SSKDLAKHIQVCDGMDLTPKIHDLKPQ 29

Db 5824 ANKQVDKQVQALIDEIDRNPNLTDEKQ 5851

RESULT 12

B86203

hypotheical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C:Accession: B86203
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizart, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, X.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: B86203
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-99 <STO>
 A:Cross-references: UNIPROT:Q9MY9; GB:AE005172; NID:g7523711; PIDN:AAF63150.1; GSPDB:GN
 C:Genetics:
 A:Map position: 1
 C:Superfamily: glutaredoxin; glutaredoxin homology

Query Match 29.9%; Score 49; DB 2; Length 99;
 Best Local Similarity 38.7%; Pred. No. 9.4;
 Matches 12; Conservative 6; Mismatches 11; Indels 2; Gaps 1;

QY 2 SSKDLAKHIQVVCVGMGLTPKIH--DLKPOC 30

DB 19 SSCCLSYAVQVLFQDLGVNPKIHEDKDPCC 49

RESULT 13

NRRT

pancreatic ribonuclease (EC 3.1.27.5) precursor - rat
 N:Alternate names: RNase 1; RNase A
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 14-Nov-1993 #sequence_revision 14-Nov-1993 #text_change 09-Jul-2004
 C:Accession: A92356; A90585; A91313; A00831
 R:MacDonald, R.J.; Stary, S.J.; Swift, G.H.
 J. Biol. Chem. 257, 14582-14585, 1982
 A:Title: Rat pancreatic ribonuclease messenger RNA. The nucleotide sequence of the entire
 A:Reference number: A92356; MUID:83082740; PMID:7174650
 A:Accession: A92356
 A:Molecule type: mRNA
 A:Residues: 1-152 <MAC>
 A:Cross-references: UNIPROT:P00684
 R:Baintema, J.J.; Gruber, M.
 Biochim. Biophys. Acta 310, 161-173, 1973
 A:Title: Rat pancreatic ribonuclease. II. Amino acid sequence.
 A:Reference number: A90585; MUID:73194545; PMID:4710592
 A:Accession: A90585
 A:Molecule type: protein
 A:Residues: 26-97, 'D', 99-123, 'T', 125, 'N', 127-128, 'N', 130, 'E', 132-152 <BEI>
 R:Baintema, J.J.
 FEBS Lett. 159, 191-195, 1983
 A:Title: Rat pancreatic ribonuclease: agreement between the corrected amino acid sequen
 A:Reference number: A91313; MUID:83262417; PMID:6873294
 A:Accession: A91313
 A:Molecule type: protein
 A:Residues: 26-152 <BE2>
 C:Superfamily: pancreatic ribonuclease
 C:Keywords: hydrolase; nucleic acid digestion; pancreas
 F:1-25/Domain: signal sequence #status predicted <SIG>
 F:26-152/Product: pancreatic ribonuclease #status experimental <MAT>
 F:40,69,147/Active site: His, Lys, His #status predicted
 F:54-112,68-123,86-138,93-100/Disulfide bonds: #status predicted

Query Match 29.9%; Score 49; DB 1; Length 152;
 Best Local Similarity 39.4%; Pred. No. 15;
 Matches 13; Conservative 2; Mismatches 8; Indels 10; Gaps 2;

QY 1 GSSK-----DLAKHIQVVCVGMGLTPKIH 24

DB 116 GSSKYPNCDYTTTDSQRHIIACDGNPVVP-VH 147

RESULT 14

F81689

conserved hypothetical protein TC0545 [imported] - Chlamydia muridarum (strain Nigg)
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 09-Jul-2004
 C:Accession: F81689
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, I.
 C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
 Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39.
 A:Reference number: A81500; MUID:20150255; PMID:10684935
 A:Accession: F81689
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-186 <TET>
 A:Cross-references: UNIPROT:Q9PKC1; GB:AE002323; GB:AE002160; NID:g7190585; PIDN:AAF39386

A:Experimental source: strain Nigg (Mopn)
 C:Genetics:
 A:Gene: TC0545

Query Match 29.9%; Score 49; DB 2; Length 186;
 Best Local Similarity 47.4%; Pred. No. 18;
 Matches 9; Conservative 5; Mismatches 1; Indels 4; Gaps 1;

QY 4 KDLAKHIQVVCVGMGLTPK 22

DB 37 RDIAKHAQI----LDMTPK 51

RESULT 15

F71534

hypothetical protein CT273 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jul-2004
 C:Accession: F71534
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
 Science 282, 734-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia tract
 A:Reference number: A71570; MUID:99008089; PMID:9784136
 A:Accession: F71534
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-188 <ARN>

A:Cross-references: UNIPROT:O84275; GB:AE001300; GB:AE001273; NID:g3328682; PIDN:AAC67866
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: CT273

Query Match 29.9%; Score 49; DB 2; Length 188;
 Best Local Similarity 47.4%; Pred. No. 18;
 Matches 9; Conservative 5; Mismatches 1; Indels 4; Gaps 1;

QY 4 KDLAKHIQVVCVGMGLTPK 22

DB 37 RDIAKHAQI----LDMTPK 51

Search completed: November 10, 2004, 12:29:29

JOB time : 7.06481 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 11:27:01 ; Search time 32.8241 Seconds
(without alignments)
525.871 Million cell updates/sec

Title: US-10-092-750-23

Perfect score: 164

Sequence: 1 GSSKDLAKHIQVVCDDGMDLTPKIHDLKPOC 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825191 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	152	92.7	458	Q91YS0	Q91YS0 mus musculus
2	152	92.7	929	1 KDGZ_MOUSE	Q80up3 mus musculus
3	152	92.7	929	1 KDGZ_RAT	Q08560 rattus norv
4	152	92.7	945	2 Q8ZT25	Q8ZT25 homo sapien
5	152	92.7	945	2 Q8ZVG7	Q8ZVG7 homo sapien
6	152	92.7	945	2 BAC85894	BAC85894 homo sapi
7	152	92.7	945	2 BAC86770	BAC86770 homo sapi
8	152	92.7	1117	1 KDGZ_HUMAN	Q13574 homo sapien
9	148	90.2	929	2 Q81VW9	Q81VW9 homo sapien
10	118	72.0	840	2 Q810C4	Q810C4 rattus norv
11	118	72.0	1050	2 Q810C5	Q810C5 rattus norv
12	118	72.0	1065	1 KDG1_HUMAN	Q75912 homo sapien
13	84	51.2	937	1 KDG1_CABEL	Q10024 caenorhabdi
14	65	39.6	842	2 Q7QFJ1	Q7QFJ1 anopheles g
15	62	37.8	332	2 Q7PCP4	Q7PCP4 anopheles g
16	58.5	35.7	503	2 Q8E8F9	Q8E8F9 shewanella
17	56	34.1	1898	2 Q8EKM6	Q8EKM6 shewanella
18	54	32.9	318	2 Q8W0F0	Q8W0F0 oryza sativ
19	54	32.9	422	2 Q8L0B9	Q8L0B9 streptomyce
20	53	32.3	548	2 Q7S6Y2	Q7S6Y2 neurospora
21	53	32.3	640	2 Q75IP7	Q75IP7 oryza sativ
22	53	32.3	640	2 AAS01975	AAS01975 oryza sat
23	52.5	32.0	152	1 RNP_RATRT	Q9Wux3 rattus ratt
24	52.5	32.0	1087	2 Q8F644	Q8F644 xanthomonas
25	52	31.7	189	2 Q9DSB7	Q9DSB7 mus musculus
26	52	31.7	262	1 ZFPF_MYCBO	Q7U0P8 mycobacteri
27	52	31.7	262	1 ZFPF_MYCTU	OS3434 mycobacteri
28	52	31.7	278	2 Q8EGX4	Q8EGX4 shewanella
29	52	31.7	279	2 Q7VMU1	Q7VMU1 haemophilus
30	52	31.7	421	2 Q8NKE1	Q8NKE1 drosophila
31	52	31.7	421	2 AAS93767	AAS93767 drosophila

RESULT 1

ID	Q91YS0	PRELIMINARY;	PRT;	458 AA.
AC	Q91YS0;			
DT	01-DEC-2001 (TrEMBLrel. 19, Created)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	Dgkz protein (Fragment).			
GN	Name=Dgkz;			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129; TISSUE=Mammary tumor. Brcal-/fl;			
RC	MEDLINE=22388257; PubMed=12477932;			
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,			
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,			
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,			
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,			
RA	Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,			
RA	Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,			
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,			
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,			
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,			
RA	Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,			
RA	Fahy J., Heiton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,			
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,			
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,			
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,			
RA	Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,			
RA	Jones S.J., Marra M.A.;			
RT	"Generation and initial analysis of more than 15,000 full-length human			
RT	and mouse cDNA sequences."			
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=129; TISSUE=Mammary tumor. Brcal-/fl;			
RA	Strausberg R.,			
RL	Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC014860; AAH14860.1; -			
DR	MGD; MGI:1278339; Dgkz.			
DR	GO; GO:0004143; F:diacylglycerol kinase activity; IEA.			
DR	GO; GO:0007205; P:protein kinase C activation; IEA.			
DR	InterPro; IPR002110; ANK.			
DR	InterPro; IPR000756; DAGK.			
DR	InterPro; IPR001206; DAGKC.			
DR	Pfam; PF00023; Ank; 2.			
DR	Pfam; PF00609; DAGK acc; 1.			
DR	PRINTS; PR01415; ANKYRIN.			
DR	ProDom; PD002939; DAGKa; 1.			
DR	SMART; SM00248; ANK; 2.			

ALIGNMENTS

Q832d4 enterococu
Q18559 caenorhabdi
Q823n7 chlamydophi
Q7v6m1 prochloroco
Q7tp58 rattus norv
Q810c3 rattus norv
O50510 streptomyce
Q7qib4 anopheles g
Q7yu71 drosophila
Q09103 drosophila
Q8sv47 drosophila
Q73wf7 mycobacteri
Aas05020 mycobacte
Q8ks50 salmonella

32 51.5 31.4 91 2 Q832D4
33 51.5 31.4 7829 2 Q18559
34 51 31.1 182 2 Q823N7
35 51 31.1 386 2 Q7V6M1
36 51 31.1 395 2 Q7TP58
37 51 31.1 591 2 Q810C3
38 51 31.1 818 2 O50510
39 51 31.1 826 2 Q7QIB4
40 51 31.1 1027 2 Q7YU71
41 51 31.1 1454 1 KDGE DROME
42 51 31.1 1457 2 Q8SV47
43 50.5 30.8 264 2 Q73WF7
44 50.5 30.8 264 2 AAS05020
45 50.5 30.8 269 2 Q8KS50

DR SMART; SM00045; DAGKA; 1.
 DR PROSITE; PS00088; ANK_REPEAT; 2.
 DR PROSITE; PS00297; ANK_REPEAT; 1.
 KW ANK repeat.
 FT NON_TER
 SQ SEQUENCE 458 AA; 51141 MW; B68D9B2B9D9B9956 CRC64;

Query Match 92.7%; Score 152; DB 2; Length 458;
 Best Local Similarity 93.3%; Pred. No. 3.5e-13;
 Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSSKDLAKHIQVCGMDLTPKIHDLKPOC 30
 DB 24 GSSKDLAKHIRVCGMDLTPKIQDLKPOC 53

RESULT 2
 ID KQZ_MOUSE STANDARD; PRT; 929 AA.

AC Q80UP3;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Diacylglycerol kinase, zeta (EC 2.7.1.107) (Diglyceride kinase) (DGK-zeta) (DAG kinase zeta).
 GN Mus musculus (Mouse).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1] _
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Breast tumor;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.B., Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loguillano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettner M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RL CC -!- FUNCTION: Displays a strong preference for 1,2-diacylglycerols over 1,3-diacylglycerols, but lacks substrate specificity among molecular species of long chain diacylglycerols (By similarity).
 CC -!- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol = ADP + 1,2-diacylglycerol 3-phosphate.
 CC -!- SUBUNIT: Interacts with the PDZ domain of the syntrophin SNTG1 (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- SIMILARITY: Belongs to the eukaryotic diacylglycerol kinase family.
 CC -!- SIMILARITY: Contains 2 ANK repeats.
 CC -!- SIMILARITY: Contains 2 zinc-dependent phorbol-ester and DAG binding domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR ENBL; BC049228; A4H49228.1; -.
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000756; DAGKA.
 DR InterPro; IPR001206; DAGKC.
 DR InterPro; IPR002219; DAG_PE-bind.
 DR Pfam; PF00609; DAGK_acc; 1.
 DR Pfam; PF00781; DAGK_cat; 1.
 DR PRINTS; PR01415; ANKYRIN.
 DR ProDom; PD002939; DAGKA; 1.
 DR ProDom; PD005043; DAGKC; 1.
 DR SMART; SM00248; ANK; 2.
 DR SMART; SM00109; CI; 2.
 DR SMART; SM00045; DAGKA; 1.
 DR SMART; SM00046; DAGKC; 1.
 DR PROSITE; PS00088; ANK_REPEAT; 2.
 DR PROSITE; PS00297; ANK_REPEAT; 1.
 DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; FALSE NEG.
 DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; FALSE NEG.
 KW ANK repeat; Kinase; Multigene family; Nuclear protein; Phorbol-ester binding; Repeat; Transferase.
 FT DOMAIN 98 153 Phorbol-ester and DAG binding 1.
 FT DOMAIN 173 231 Phorbol-ester and DAG binding 2.
 FT DOMAIN 294 420 Catalytic-A (Potential).
 FT DOMAIN 447 604 Catalytic-B (Potential).
 FT REPEAT 823 853 ANK 1.
 FT REPEAT 858 887 ANK 2.
 FT DOMAIN 67 72 Poly-pro.
 FT DOMAIN 260 263 Poly-lys.
 FT DOMAIN 372 375 Poly-pro.
 SQ SEQUENCE 929 AA; 104046 MW; AC3E46790E223583 CRC64;

Query Match 92.7%; Score 152; DB 1; Length 929;
 Best Local Similarity 93.3%; Pred. No. 7.5e-13;
 Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSSKDLAKHIQVCGMDLTPKIHDLKPOC 30
 DB 496 GSSKDLAKHIRVCGMDLTPKIQDLKPOC 525

RESULT 3
 KQZ_MOUSE STANDARD; PRT; 929 AA.

AC Q08560;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Diacylglycerol kinase, zeta (EC 2.7.1.107) (Diglyceride kinase) (DGK-zeta) (DAG kinase zeta) (DGK-IV) (104 kDa diacylglycerol kinase).
 GN Name=Dgkz; Synonyms=Dagk6;
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1] _
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wistar;
 RX MEDLINE=97008156; PubMed=8855332;
 RA Goto K., Kondo H.; "A 104-kDa diacylglycerol kinase containing ankyrin-like repeats localizes in the cell nucleus.";
 RT Proc. Natl. Acad. Sci. U.S.A. 93:11196-11201 (1996).
 RL CC -!- FUNCTION: Displays a strong preference for 1,2-diacylglycerols over 1,3-diacylglycerols, but lacks substrate specificity among molecular species of long chain diacylglycerols.
 CC -!- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol = ADP + 1,2-diacylglycerol 3-phosphate.
 CC -!- SUBUNIT: Interacts with the PDZ domain of the syntrophin SNTG1 (By similarity).

```

CC CC -!- SUBCELLULAR LOCATION: Nuclear.
CC CC -!- SIMILARITY: Belongs to the eukaryotic diacylglycerol kinase
CC CC family.
CC CC -!- SIMILARITY: Contains 2 ANK repeats.
CC CC -!- SIMILARITY: Contains 2 zinc-dependent phosphol-ester and DAG
CC CC binding domains.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use. By non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; D78588; BAA18942.1; -
CC CC FIR; JC6124; JC6124.
CC CC RGD; 70929; Dgk2.
CC CC InterPro; IPR002110; ANK.
CC CC InterPro; IPR000756; DAGKA.
CC CC InterPro; IPR001206; DAGKC.
CC CC InterPro; IPR002219; DAG_PE-bind.
CC CC Pfam; PF00023; Ank; 2.
CC CC Pfam; PF00609; DAGK_acc; 1.
CC CC Pfam; PF00781; DAGK_cat; 1.
CC CC PRINTS; PR01415; ANKYRIN.
CC CC ProDom; PD002939; DAGKA; 1.
CC CC ProDom; PD005043; DAGKC; 1.
CC CC SMART; SM00248; ANK; 2.
CC CC SMART; SM00109; Cl; 2.
CC CC SMART; SM00045; DAGK; 1.
CC CC SMART; SM00046; DAGKC; 1.
CC CC PROSITE; PS50088; ANK_REPEAT; 2.
CC CC PROSITE; PS50297; ANK_REPEAT; 1.
CC CC PROSITE; PS50479; DAG_PE_BIND_DOM_1; FALSE_NEG.
CC CC PROSITE; PS50081; DAG_PE_BIND_DOM_2; FALSE_NEG.
CC CC ANK repeat; Kinase; Multigene family; Nuclear protein;
CC CC Phorbol-ester binding; Repeat; Transferase.
CC CC FT DOMAIN 98 153 Phorbol-ester and DAG binding 1.
CC CC FT DOMAIN 173 231 Phorbol-ester and DAG binding 2.
CC CC FT DOMAIN 294 420 Catalytic-A (Potential).
CC CC FT DOMAIN 447 604 Catalytic-B (Potential).
CC CC FT REPEAT 823 853 ANK 1.
CC CC FT REPEAT 858 887 ANK 2.
CC CC FT DOMAIN 67 72 Poly-Pro.
CC CC FT DOMAIN 260 263 Poly-Lys.
CC CC FT DOMAIN 372 375 Poly-Pro.
CC CC FT SEQUENCE 929 AA; 104009 MW; 6279872004F5F38D CRC64;
CC CC -----
CC CC Query Match 92.7%; Score 152; DB 1; Length 929;
CC CC Best Local Similarity 93.3%; Pred. No. 7.5e-13;
CC CC Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC CC -----
CC CC QY 1 GSSKDLAKHIQVVCDDMDLTPKIHDLKPC 30
CC CC Db 496 GSSKDLAKHIRVVCDDMDLTPKIQLKPC 525
CC CC -----
CC CC RESULT 4
CC CC Q6ZT25 PRELIMINARY; PRT; 945 AA.
CC CC AC Q6ZT25;
CC CC DT 05-JUL-2004 (TrEMBLrel. 27, Created)
CC CC DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
CC CC DE Hypothetical protein FLJ45023.
CC CC OS Homo sapiens (Human).
CC CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC CC OX NCBI_TaxID=9606;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RC TISSUE=Brain;

```

```

RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Orsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuna M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Suzuki Y., Sugano S., Nagahori K., Masuho Y., Nagai K., Isogai T.;
RA Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK126970; BAC86770.1; -
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000756; DAGKA.
DR InterPro; IPR001206; DAGKC.
DR InterPro; IPR002219; DAG_PE-bind.
DR Pfam; PF00023; Ank; 2.
DR Pfam; PF00609; DAGK_acc; 1.
DR Pfam; PF00781; DAGK_cat; 1.
DR PRINTS; PR01415; ANKYRIN.
DR ProDom; PD002939; DAGKA; 1.
DR ProDom; PD005043; DAGKC; 1.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00109; Cl; 2.
DR SMART; SM00045; DAGKA; 1.
DR SMART; SM00046; DAGKC; 1.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REPEAT; 1.
KW ANK repeat; Kinase.
SQ SEQUENCE 945 AA; 106082 MW; 4DAEA0EBD83D021 CRC64;
CC -----
CC Query Match 92.7%; Score 152; DB 2; Length 945;
CC Best Local Similarity 93.3%; Pred. No. 7.6e-13;
CC Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC -----
CC CC QY 1 GSSKDLAKHIQVVCDDMDLTPKIHDLKPC 30
CC CC Db 512 GSSKDLAKHIRVVCDDMDLTPKIQLKPC 541
CC CC -----
CC CC RESULT 5
CC CC Q6ZV77 PRELIMINARY; PRT; 945 AA.
CC CC AC Q6ZV77;
CC CC DT 05-JUL-2004 (TrEMBLrel. 27, Created)
CC CC DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
CC CC DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
CC CC DE Hypothetical protein FLJ42603.
CC CC OS Homo sapiens (Human).
CC CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC CC OX NCBI_TaxID=9606;
CC CC RN [1]
CC CC RP SEQUENCE FROM N.A.
CC CC RC TISSUE=Cerebellum;
CC CC RA Ninomiya K., Wagatsuna M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
CC CC Katsuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
CC CC Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
CC CC Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
CC CC Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
CC CC Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
CC CC Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
CC CC Suzuki Y., Sugano S., Nagahori K., Masuho Y., Nagai K., Isogai T.;
CC CC Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR GO; GO:0016301; F:kinase activity; IEA.
DR EMBL; AK124594; BAC85894.1; -
DR GO; GO:0016301; F:kinase activity; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000756; DAGKA.
DR InterPro; IPR001206; DAGKC.
DR InterPro; IPR002219; DAG_PE-bind.
DR Pfam; PF00023; Ank; 2.
DR Pfam; PF00609; DAGK_acc; 1.
DR Pfam; PF00781; DAGK_cat; 1.
DR PRINTS; PR01415; ANKYRIN.
DR ProDom; PD002939; DAGKA; 1.
DR ProDom; PD005043; DAGKC; 1.

```

DR SMART; SM00248; ANK; 2.
DR SMART; SM00109; C1; 2.
DR SMART; SM00045; DAGKA; 1.
DR SMART; SM00046; DAGKC; 1.
DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REPEAT; 1.
KW ANK repeat, Kinase.
SQ SEQUENCE 945 AA; 106030 MW; 6EF612B9FE91F7E7 CRC64;

Query Match 92.7%; Score 152; DB 2; Length 945;
Best Local Similarity 93.3%; Pred. No. 7.6e-13;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSSKDLAKHIQVVCMDLTPKIHDLKPOC 30
|||||:|||||:|||||:|||||
DB 512 GSSKDLAKHIRVVCDGMDLTPKIQDLKPOC 541

RESULT 6

BAC85894 PRELIMINARY; PRT; 945 AA.

ID BAC85894
AC BAC85894
DT 02-MAR-2004 (TREMBLrel. 27, Created)
DT 02-MAR-2004 (TREMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBLrel. 27, Last annotation update)
DE CDNA FLJ42603 fis. clone BRACE3011421, highly similar to
DE Diacylglycerol kinase, zeta (EC 2.7.1.107).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primata; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RA Ninomiya K., Hagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AX124594; BAC85894.1; -.
KW Kinase.

SQ SEQUENCE 945 AA; 106030 MW; 6EF612B9FE91F7E7 CRC64;

Query Match 92.7%; Score 152; DB 2; Length 945;
Best Local Similarity 93.3%; Pred. No. 7.6e-13;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSSKDLAKHIQVVCMDLTPKIHDLKPOC 30
|||||:|||||:|||||:|||||
DB 512 GSSKDLAKHIRVVCDGMDLTPKIQDLKPOC 541

RESULT 7

BAC86770 PRELIMINARY; PRT; 945 AA.

ID BAC86770
AC BAC86770
DT 02-MAR-2004 (TREMBLrel. 27, Created)
DT 02-MAR-2004 (TREMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBLrel. 27, Last annotation update)
DE CDNA FLJ45023 fis. clone BRAH3017180, highly similar to
DE Diacylglycerol kinase, zeta (EC 2.7.1.107).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;

RA Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA Suzuki Y., Sugano S., Nagahara K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK126970; BAC86770.1; -.
KW Kinase.

SQ SEQUENCE 945 AA; 106082 MW; 4DAEA0EB83D021 CRC64;

Query Match 92.7%; Score 152; DB 2; Length 945;

Best Local Similarity 93.3%; Pred. No. 7.6e-13;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GSSKDLAKHIQVVCMDLTPKIHDLKPOC 30
|||||:|||||:|||||:|||||
DB 512 GSSKDLAKHIRVVCDGMDLTPKIQDLKPOC 541

RESULT 8

KDGZ HUMAN STANDARD; PRT; 1117 AA.

ID KDGZ HUMAN
AC Q13574; O00542;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-OCT-2001 (Rel. 40, Last sequence update)
DT 03-JUN-2004 (Rel. 44, Last annotation update)
DE Diacylglycerol kinase, zeta (EC 2.7.1.107) (Diglyceride kinase) (DGK-
DE zeta) (DAG kinase zeta).
GN Name=DGKZ; Synonyms=DAGK6;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM SHORT).
RC TISSUE=Endothelial cells;
RX MEDLINE=96215319; PubMed=8626588;
RA Bunting M., Tang W., Zimmerman G.A., McIntyre T.M., Prescott S.M.;
RX "Molecular cloning and characterization of a novel human
RT diacylglycerol kinase zeta.";
RL J. Biol. Chem. 271:10230-10236(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM LONG).
RC TISSUE=Skeletal muscle;
RX MEDLINE=97303161; PubMed=9159104;
RA Ding L., Bunting M., Topham M.K., McIntyre T.M., Zimmerman G.A.,
RA Prescott S.M.;
RT "Alternative splicing of the human diacylglycerol kinase zeta gene in
RL muscle.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:5519-5524(1997).
RN [3]
RP PHOSPHORYLATION, AND SUBCELLULAR LOCATION.
RX MEDLINE=98379933; PubMed=9716136; DOI=10.1038/29337;
RA Topham M.K., Bunting M., Zimmerman G.A., McIntyre T.M.,
RA Blackshear P.J., Prescott S.M.;
RT "Protein kinase C regulates the nuclear localization of diacylglycerol
RT kinase-zeta.";
RL Nature 394:697-700(1998).
RN [4]
RP INTERACTION WITH SNTG1, AND MUTAGENESIS OF 1115-THR-ALA-1116.
RX MEDLINE=21336641; PubMed=11352924; DOI=10.1074/jbc.M104156200;
RA Hogan A., Shepherd L., Chabot J., Quenneville S., Prescott S.M.,
RA Topham M.K., Gee S.H.;
RT "Interaction of gamma 1-syntrophin with diacylglycerol kinase-zeta.
RT Regulation of nuclear localization by PDZ interactions.";
RL J. Biol. Chem. 276:26526-26533(2001).
CC -I- FUNCTION: Displays a strong preference for 1,2-diacylglycerols
CC over 1,3-diacylglycerols, but lacks substrate specificity among
CC molecular species of long chain diacylglycerols.
CC -I- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol = ADP + 1,2-
CC diacylglycerol 3-phosphate.

```
CC -!- SUBUNIT: Interacts with the PDZ domain of the syntrophin SNTG1.
CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q13574-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q13574-2; Sequence=VSP_001268;
CC -!- TISSUE SPECIFICITY: Highest levels in brain, and substantial
CC levels in skeletal muscle, heart, and pancreas.
CC -!- PTM: Phosphorylation of the MARCKS homology domain by PKC reduces
CC nuclear accumulation of DGK-zeta.
CC -!- SIMILARITY: Belongs to the eukaryotic diacylglycerol kinase
CC family.
CC -!- SIMILARITY: Contains 2 ANK repeats.
CC -!- SIMILARITY: Contains 2 zinc-dependent phorbol-ester and DAG
CC binding domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U51477; AAC50478.1; -.
CC EMBL; U94905; AAB60859.1; -.
CC Genew; HGNC:2857; DGKZ.
CC MIM; 601441; -.
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0005524; P:APP binding; TAS.
CC GO; GO:0004143; P:diacylglycerol kinase activity; TAS.
CC GO; GO:0007165; P:signal transduction; TAS.
CC InterPro; IPR002110; ANK.
CC InterPro; IPR007566; DAGKA.
CC InterPro; IPR001206; DAGK.
CC InterPro; IPR002219; DAG_PE-bind.
CC Pfam; PF00023; Ank; 2.
CC Pfam; PF00609; DAGK_acc; 1.
CC Pfam; PF00781; DAGK_cat; 1.
CC PRINTS; PR01415; ANKYRIN.
CC ProDom; PD002939; DAGKA; 1.
CC ProDom; PD005043; DAGKC; 1.
CC SMART; SM00248; ANK; 2.
CC SMART; SM00109; Cl; 2.
CC SMART; SM00045; DAGKA; 1.
CC PROSITE; PS50088; ANK_REPEAT; 2.
CC PROSITE; PS50297; ANK_REPEAT; 1.
CC PROSITE; PS50479; DAG_PE_BIND_DOM_1; FALSE_NEG.
CC PROSITE; PS50081; DAG_PE_BIND_DOM_2; FALSE_NEG.
CC Alternative splicing; ANK repeat; Kinase; Multigene family;
CC Nuclear protein; Phorbol-ester binding; Phosphorylation; Repeat;
CC Transferrase.
CC -----
CC DOMAIN 287 341 Phorbol-ester and DAG binding 1.
CC FT DOMAIN 361 419 Phorbol-ester and DAG binding 2.
CC FT DOMAIN 448 462 MARCKS HOMOLGY.
CC FT DOMAIN 482 608 Catalytic-A (Potential).
CC FT DOMAIN 635 792 Catalytic-B (Potential).
CC FT REPEAT 1011 1041 ANK 1.
CC FT REPEAT 1046 1075 ANK 2.
CC FT DOMAIN 256 261 Poly-Pro.
CC FT DOMAIN 448 451 Poly-Lys.
CC FT DOMAIN 560 563 Poly-Pro.
CC FT VARSPIC 1 243 ASSSLAQRSSAQLGGLSCGVRQSSRRASSTVPPSC
CC METFRHFGKVPGEQSRPSSVGLPTGKARRRSPAGO
CC NRRFIVDVLPTQPTTVGAQLGLLTLGLVGNNEEGVQ
CC EDVVAESAIAIQPGTKFTGPPPPRGAQLPLPRYVRAS
CC HCCPADAVYDALWGLHYRLSRPSGQHPGPGRRAS
CC CTGATGMLPTVRVLRQRQVLRKAAGPQAWALLA ->
CC MEPRDGSPEARSSDSASASSGSRDAGEPDKAPRLN
CC KERFGLRFGHR (in isoform Short).
```

```
FT MUTAGEN 1115 1116 /FTId=VSP_001268.
FT TA->NS: Loss of interaction with SNTG1.
SQ SEQUENCE 1117 AA; 124122 MW; 213BC8ADBD4E1402 CRC64;
Query Match: 92.7%; Score 152; DB 1; Length 1117;
Best Local Similarity 93.3%; Pred. NO. 9,1e-13;
Matches 28; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 GSSKDLAKHIQVVCVCGMDLTPKIHDLKPOC 30
Db 684 GSSKDLAKHIVVVCVCGMDLTPKIQDLKPOC 713
RESULT 9
Q81VW9 PRELIMINARY; PRT; 929 AA.
AC Q81VW9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Diacylglycerol kinase, zeta 10kDa, isoform 2.
GN Name=DGKZ;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Sanchez A.,
RA Paney J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RA "Generation and initial analysis of more than 15,000 full-length human
RA and mouse cDNA sequences.";
RA Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC041770; AAH41770.1; -.
DR GO; GO:0004143; F:diacylglycerol kinase activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0007205; P:protein kinase C activation; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000756; DAGKA.
DR InterPro; IPR001206; DAGK.
DR Pfam; PF00023; Ank; 2.
DR Pfam; PF00609; DAGK_acc; 1.
DR Pfam; PF00781; DAGK_cat; 1.
DR PRINTS; PR01415; ANKYRIN.
DR ProDom; PD002939; DAGKA; 1.
DR ProDom; PD005043; DAGKC; 1.
DR SMART; SM00248; ANK; 2.
DR SMART; SM00109; Cl; 2.
DR SMART; SM00045; DAGKA; 1.
DR SMART; SM00046; DAGKC; 1.
```

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF061936; AACG2010.1; -.
EMBL; AF219939; AAF43006.1; JOINED.
EMBL; AF219907; AAF43006.1; JOINED.
EMBL; AF219908; AAF43006.1; JOINED.
EMBL; AF219909; AAF43006.1; JOINED.
EMBL; AF219910; AAF43006.1; JOINED.
EMBL; AF219911; AAF43006.1; JOINED.
EMBL; AF219912; AAF43006.1; JOINED.
EMBL; AF219913; AAF43006.1; JOINED.
EMBL; AF219914; AAF43006.1; JOINED.
EMBL; AF219915; AAF43006.1; JOINED.
EMBL; AF219916; AAF43006.1; JOINED.
EMBL; AF219917; AAF43006.1; JOINED.
EMBL; AF219918; AAF43006.1; JOINED.
EMBL; AF219919; AAF43006.1; JOINED.
EMBL; AF219920; AAF43006.1; JOINED.
EMBL; AF219921; AAF43006.1; JOINED.
EMBL; AF219922; AAF43006.1; JOINED.
EMBL; AF219923; AAF43006.1; JOINED.
EMBL; AF219924; AAF43006.1; JOINED.
EMBL; AF219925; AAF43006.1; JOINED.
EMBL; AF219926; AAF43006.1; JOINED.
EMBL; AF219927; AAF43006.1; JOINED.
EMBL; AF219928; AAF43006.1; JOINED.
EMBL; AF219929; AAF43006.1; JOINED.
EMBL; AF219930; AAF43006.1; JOINED.
EMBL; AF219931; AAF43006.1; JOINED.
EMBL; AF219932; AAF43006.1; JOINED.
EMBL; AF219933; AAF43006.1; JOINED.
EMBL; AF219934; AAF43006.1; JOINED.
EMBL; AF219935; AAF43006.1; JOINED.
EMBL; AF219936; AAF43006.1; JOINED.
EMBL; AF219937; AAF43006.1; JOINED.
EMBL; AF219938; AAF43006.1; JOINED.
EMBL; AF219939; AAF43006.1; JOINED.
MIM; 604072; -.
GO; GO:0005737; C:cytoplasm; TAS.
GO; GO:0005634; C:nucleus; TAS.
GO; GO:0004143; F:diacylglycerol kinase activity; TAS.
InterPro; IPR002110; ANK.
InterPro; IPR000756; DAGKa.
InterPro; IPR001206; DAGKc.
InterPro; IPR002219; DAG PE-bind.
Pfam; PF00023; ANK; 2.
Pfam; PF00609; DAGK acc; 1.
Pfam; PF00781; DAGK_cat; 1.
PRINTS; PR01415; ANKXRN.
ProDom; PD002359; DAGKa; 1.
ProDom; PD005043; DAGKc; 1.
SMART; SM00248; ANK; 2.
SMART; SM00109; C1; 2.
SMART; SM00045; DAGKa; 1.
SMART; SM00046; DAGKc; 1.
PROSITE; PS50088; ANK_REPEAT; 2.
PROSITE; PS50297; ANK_REPEAT_REGION; 1.
PROSITE; PS50479; DAG PE BIND DOM 1; FALSE NEG.
PROSITE; PS50081; DAG PE BIND DOM 2; FALSE NEG.
KW ANK repeat; Kinase; Multigene family; Nuclear protein; Polymorphism;
Repeat; Transferase.
FT DOMAIN 178 232 Phorbol-ester and DAG binding 1.
FT DOMAIN 231 309 Phorbol-ester and DAG binding 2.
FT DOMAIN 374 500 Catalytic-A (Potential).
FT DOMAIN 526 683 Catalytic-B (Potential).
FT REPEAT 958 990 ANK 1.

FT REPEAT 997 1026 ANK 2.
FT DOMAIN 20 31 Poly-Ala.
FT DOMAIN 69 74 Poly-Ser.
FT DOMAIN 95 102 Poly-Ala.
FT VARIANT 153 153 L -> F.
FT /FTID=VAR 010190.
FT CONFLICT 160 160 A -> P (in Ref. 2).
SQ SEQUENCE 1065 AA; 116996 MW; B84971AA7630A799 CRC64;

Query Match 72.0%; Score 118; DB 1; Length 1065;
Best Local Similarity 72.4%; Pred. No. 6.8e-08;
Matches 21; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 2 SSKDLAKHIQVCDGMDLTPKIHDLKPOC 30
|||:||||:||||:||||:||||:||||:||||:
Db 576 SSRDLKHKVXVCDGTDLPKIQELAFQC 604

RESULT 13
KOGM_CABEL STANDARD; PRT; 937 AA.
ID KOGM_CABEL Q10024; Q09588;
DT 01-OCT-1996 (Rel. 34, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Putative diacylglycerol kinase K06A1.6 (EC 2.7.1.107) (Diglyceride
kinase) (DGK) (DAG kinase).
GN ORFNames=K06A1.6, T28D9.5;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Fulton L.; (APR-1995) to the EMBL/GenBank/DBJ databases.
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RP REVISIONS.
RA Waterston R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + 1,2-diacylglycerol = ADP + 1,2-
diacylglycerol 3-phosphate.
CC -1- SIMILARITY: Belongs to the eukaryotic diacylglycerol kinase
family.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; U23449; AAC24300.2; -.
PIR; T16947; T16947.
WormPep; K06A1.6; CE28048.
InterPro; IPR000756; DAGKa.
InterPro; IPR001206; DAGKc.
Pfam; PF00609; DAGK acc; 1.
Pfam; PF00781; DAGK_cat; 1.
ProDom; PD002939; DAGKa; 1.
ProDom; PD005043; DAGKc; 1.
SMART; SM00045; DAGKa; 1.
SMART; SM00046; DAGKc; 1.
HYPOTHETICAL protein; Kinase; Transferase.
FT DOMAIN 95 157 Ser-rich.
FT DOMAIN 418 545 Catalytic-A (Potential).
FT DOMAIN 572 725 Catalytic-B (Potential).
SQ SEQUENCE 937 AA; 105197 MW; 7F9FAL63AF783790 CRC64;

Query Match 51.2%; Score 84; DB 1; Length 937;
Best Local Similarity 51.7%; Pred. No. 0.0047;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:32:37 ; Search time 11.399 Seconds
(without alignments)
191.991 Million cell updates/sec

Title: US-10-092-750-24
Perfect score: 155
Sequence: 1 GFLAAEQDIREIRKVVQSLQETAREVLTLLQG 33

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65	41.9	92	4	US-09-513-999C-6251
2	59	38.1	74	4	Sequence 6251, Ap
3	59	38.1	74	4	Sequence 38906, A
4	59	38.1	74	4	Sequence 54123, A
5	57	36.8	734	4	Sequence 32311, A
6	56	36.1	760	4	Sequence 30703, A
7	56	36.1	862	4	Sequence 31724, A
8	53	34.2	294	4	Sequence 14, Appl
9	53	34.2	410	4	Sequence 13944, A
10	53	34.2	430	4	Sequence 7, Appl
11	51	32.9	2854	4	Sequence 6, Appl
12	50.5	32.6	963	4	Sequence 1, Appl
13	50.5	32.6	963	4	Sequence 20, Appl
14	50.5	32.6	963	4	Sequence 22, Appl
15	50	32.3	129	4	Sequence 1060, Ap
16	50	32.3	640	4	Sequence 12730, A
17	50	32.3	784	4	Sequence 8058, Ap
18	50	32.3	865	3	Sequence 1, Appl
19	50	32.3	865	4	Sequence 19, Appl
20	50	32.3	865	4	Sequence 19, Appl
21	49	31.6	166	3	Sequence 19, Appl
22	49	31.6	166	3	Sequence 9, Appl
23	49	31.6	166	3	Sequence 20, Appl
24	49	31.6	166	3	Sequence 21, Appl
25	49	31.6	166	4	Sequence 2, Appl
26	49	31.6	166	4	Sequence 1219, Ap
27	49	31.6	166	5	Sequence 9, Appl

28	49	31.6	206	4	US-09-252-991A-17940	Sequence 17940, A
29	49	31.6	299	4	US-09-270-767-37559	Sequence 37559, A
30	49	31.6	299	4	US-09-270-767-52776	Sequence 52776, A
31	49	31.6	318	4	US-08-555-755C-6	Sequence 6, Appl
32	49	31.6	423	4	US-09-270-767-46381	Sequence 46381, A
33	49	31.6	1318	4	US-09-540-236-3623	Sequence 3623, Ap
34	48	31.0	211	3	US-09-323-872A-12	Sequence 12, Appl
35	48	31.0	211	4	US-09-072-433-21	Sequence 21, Appl
36	48	31.0	273	4	US-09-578-063-6	Sequence 6, Appl
37	48	31.0	456	4	US-09-578-063-5	Sequence 5, Appl
38	48	31.0	458	3	US-09-039-609-2	Sequence 2, Appl
39	48	31.0	475	3	US-09-578-063-3	Sequence 3, Appl
40	48	31.0	493	4	US-09-134-000C-4034	Sequence 4034, Ap
41	48	31.0	513	1	US-08-403-866-1	Sequence 1, Appl
42	48	31.0	636	4	US-09-328-353-5495	Sequence 5495, Ap
43	48	31.0	752	3	US-09-817-180-2	Sequence 2, Appl
44	48	31.0	752	4	US-10-003-295-2	Sequence 2, Appl
45	48	31.0	822	3	US-09-817-180-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-09-513-999C-6251
; Sequence 6251, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 6251
; LENGTH: 92
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 24
; OTHER INFORMATION: Xaa-Ile or Val
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 28
; OTHER INFORMATION: Xaa= * or Cys or Trp
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 52
; OTHER INFORMATION: Xaa= * or Ser
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 64
; OTHER INFORMATION: Xaa-Ser or Thr
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 76
; OTHER INFORMATION: Xaa-Ala or Glu or Gly or Val
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 86
; OTHER INFORMATION: Xaa-Lys or Thr
; US-09-513-999C-6251

Query Match 41.9%; Score 65; DB 4; Length 92;
Best Local Similarity 86.7%; Pred. No. 0.096;
Matches 13; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 GLAAEQDIREIRK 15
Db 12 GLAAEQDIREEXK 26

RESULT 2
US-09-270-767-38906
; Sequence 38906, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 38906
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-38906

Query Match 38.1%; Score 59; DB 4; Length 74;
Best Local Similarity 38.7%; Pred. No. 0.49;
Matches 12; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 2 FLAAEQDIREIRKVQSLQEQTAREVLTLLQ 32
Db 3 YIDNEQEVRENIRIVVREIHLKSQAQIKLQ 33

RESULT 3
US-09-270-767-54123
; Sequence 54123, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 54123
; LENGTH: 74
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-54123

Query Match 38.1%; Score 59; DB 4; Length 74;
Best Local Similarity 38.7%; Pred. No. 0.49;
Matches 12; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 2 FLAAEQDIREIRKVQSLQEQTAREVLTLLQ 32
Db 3 YIDNEQEVRENIRIVVREIHLKSQAQIKLQ 33

RESULT 4
US-09-270-767-32311
; Sequence 32311, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 32311

Query Match 38.1%; Score 59; DB 4; Length 74;
Best Local Similarity 38.7%; Pred. No. 0.49;
Matches 12; Conservative 8; Mismatches 11; Indels 0; Gaps 0;

QY 2 FLAAEQDIREIRKVQSLQEQTAREVLTLLQ 32
Db 3 YIDNEQEVRENIRIVVREIHLKSQAQIKLQ 33

RESULT 5
US-09-252-991A-30703
; Sequence 30703, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30703
; LENGTH: 734
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30703

Query Match 36.8%; Score 57; DB 4; Length 734;
Best Local Similarity 28.6%; Pred. No. 14;
Matches 12; Conservative 10; Mismatches 10; Indels 10; Gaps 1;

QY 1 GLAAEQDIR-----EIRKVQSLQEQTAREVLTLLQ 32
Db 598 GFAWADEVRLNAQKTQATEBIQSMIQLOQGTDRDVVKVMQ 639

RESULT 6
US-09-252-991A-31724
; Sequence 31724, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31724
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31724

Query Match 36.1%; Score 56; DB 4; Length 760;
Best Local Similarity 40.7%; Pred. No. 21;
Matches 11; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 7 QDIREIRKVQSLQEQTAREVLTLLQ 33
Db 7 QDIREIRKVQSLQEQTAREVLTLLQ 33

Query Match 36.1%; Score 56; DB 4; Length 760;
Best Local Similarity 40.7%; Pred. No. 21;
Matches 11; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY 7 QDIREIRKVQSLQEQTAREVLTLLQ 33
Db 7 QDIREIRKVQSLQEQTAREVLTLLQ 33
```

```
Db      640 QDSVEQIRGVIEGLQGQTRDVVDVAVHG 666

RESULT 7
US-09-665-479A-14
; Sequence 14, Application US/09665479A
; Patent No. 6673570
; GENERAL INFORMATION:
; APPLICANT: Itoh, Fumiko
; APPLICANT: Itoh, Susumu
; APPLICANT: Heldin, Carl-Henrik
; APPLICANT: ten-Dijke, Peter
; TITLE OF INVENTION: SMAD ASSOCIATING POLYPEPTIDES
; FILE REFERENCE: L00461.70096.US
; CURRENT APPLICATION NUMBER: US/09/665,479A
; CURRENT FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: US 60/154,846
; PRIOR FILING DATE: 1999-09-20
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-665-479A-14

Query Match      36.1%; Score 56; DB 4; Length 862;
Best Local Similarity 33.3%; Pred. No. 24;
Matches 13; Conservative 11; Mismatches 9; Indels 6; Gaps 1;

Qy      1  GFLAAEQDIRREIRKVVQSLEQTAREVLTLL-----LQ 33
Db      616 GLSQAKRDVQEQMAVLMSREQVSELVRLQKNDNSLQ 654

RESULT 8
US-09-489-039A-13944
; Sequence 13944, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13944
; LENGTH: 294
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13944

Query Match      34.2%; Score 53; DB 4; Length 294;
Best Local Similarity 37.5%; Pred. No. 17;
Matches 12; Conservative 7; Mismatches 13; Indels 0; Gaps 0;

Qy      1  GFLAAEQDIRREIRKVVQSLEQTAREVLTLLQ 32
Db      263 GVHAGQNTAREMIVLPDLQAKARELTLLR 294

RESULT 9
US-09-702-953B-7
; Sequence 7, Application US/09702953B
; Patent No. 6673897
; GENERAL INFORMATION:
; APPLICANT: VLAAMS INTERUNIVERSITAIR INSTITUUT VOOR BIOTECHNOL
; TITLE OF INVENTION: NOVEL INHIBITORS OF NF-kappaB ACTIVATION
; FILE REFERENCE: 2676-4554US
; CURRENT APPLICATION NUMBER: US/09/702,953B
```

```
; CURRENT FILING DATE: 2000-10-31
; PRIOR APPLICATION NUMBER: PCT/BE99/00055
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: 98201472.2
; PRIOR FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 410
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-702-953B-7

Query Match      34.2%; Score 53; DB 4; Length 410;
Best Local Similarity 32.1%; Pred. No. 25;
Matches 9; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

Qy      4  AAEQDIRREIRKVVQSLEQTAREVLTLL 31
Db      66  ASETQLRQEVERLTERLEEKEREMQQLM 93

RESULT 10
US-09-702-953B-6
; Sequence 6, Application US/09702953B
; Patent No. 6673897
; GENERAL INFORMATION:
; APPLICANT: VLAAMS INTERUNIVERSITAIR INSTITUUT VOOR BIOTECHNOL
; TITLE OF INVENTION: NOVEL INHIBITORS OF NF-kappaB ACTIVATION
; FILE REFERENCE: 2676-4554US
; CURRENT APPLICATION NUMBER: US/09/702,953B
; CURRENT FILING DATE: 2000-10-31
; PRIOR APPLICATION NUMBER: PCT/BE99/00055
; PRIOR FILING DATE: 1999-05-05
; PRIOR APPLICATION NUMBER: 98201472.2
; PRIOR FILING DATE: 1998-05-06
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-702-953B-6

Query Match      34.2%; Score 53; DB 4; Length 430;
Best Local Similarity 32.1%; Pred. No. 27;
Matches 9; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

Qy      4  AAEQDIRREIRKVVQSLEQTAREVLTLL 31
Db      86  ASETQLRQEVERLTERLEEKEREMQQLM 113

RESULT 11
US-09-150-867-1
; Sequence 1, Application US/09150867
; Patent No. 6645748
; GENERAL INFORMATION:
; APPLICANT: Wood, Kenneth W.
; APPLICANT: Sakowicz, Roman
; APPLICANT: Goldstein, Lawrence S.B.
; APPLICANT: Cleveland, Don W.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Plus End-Directed Microtubule Motor Required for
; FILE REFERENCE: 18557C-000110US
; CURRENT APPLICATION NUMBER: US/09/150,867
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: US 60/058,645
; EARLIER FILING DATE: 1997-09-11
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
```


	Matches	10;	Conservative	4;	Mismatches	8;	Indels	0;	Gaps	0;
Qy	6	EQDIREERKVVQSLEQTAREV 27								
	:	:	:	:	:	:	:	:	:	:
Db	99	QOALHQRLETVVQOLEQVQREV 120								

Search completed: November 10, 2004, 13:43:59
Job time : 12.399 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 16:36:12 ; Search time 35.4508 Seconds
(without alignments)
328.807 Million cell updates/sec

Title: US-10-092-750-24
Perfect score: 155
Sequence: 1 GFLAAEQDIREIRKVVQSLQETAREVLTLLQ 33

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 156620 seqs, 35325886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	155	100.0	33	14	US-10-092-750-24
2	154	99.4	253	9	US-09-925-297-499
3	58	37.4	429	15	US-10-108-260A-3640
4	58	37.4	513	14	US-10-156-761-13249
5	57	36.8	234	16	US-10-437-963-106139
6	56	36.1	751	16	US-10-476-924-22
7	56	36.1	862	14	US-10-177-293-391
8	56	36.1	862	16	US-10-408-765A-2447
9	54.5	35.2	551	15	US-10-411-066-40
10	54	34.8	217	17	US-10-425-115-236205
11	54	34.8	221	17	US-10-425-115-357714
12	54	34.8	230	15	US-10-425-114-60374
13	54	34.8	230	16	US-10-767-701-43357

14	54	34.8	230	17	US-10-425-115-357715
15	54	34.8	239	15	US-10-425-114-59221
16	53.5	34.5	1177	14	US-10-128-714-3493
17	53.5	34.5	1179	14	US-10-128-714-8493
18	53	34.2	213	15	US-10-424-599-244430
19	53	34.2	242	15	US-10-282-122A-49985
20	53	34.2	250	17	US-10-425-115-361021
21	53	34.2	410	15	US-10-680-998-7
22	53	34.2	430	15	US-10-680-998-6
23	53	34.2	494	14	US-10-035-343-4
24	53	34.2	989	15	US-10-320-787-3322
25	52.5	33.9	63	16	US-10-767-701-34531
26	52.5	33.9	222	17	US-10-425-115-288907
27	52.5	33.9	226	15	US-10-425-114-63666
28	52	33.5	232	15	US-10-424-599-219142
29	52	33.5	799	15	US-10-282-122A-61690
30	52	33.5	1145	14	US-10-369-493-20484
31	52	33.5	1941	15	US-10-188-186-96
32	51	32.9	95	16	US-10-767-701-45197
33	51	32.9	162	15	US-10-424-599-160703
34	51	32.9	351	15	US-10-424-599-204556
35	51	32.9	468	14	US-10-369-493-12643
36	51	32.9	605	15	US-10-282-122A-45415
37	51	32.9	790	15	US-10-282-122A-63670
38	51	32.9	867	15	US-10-425-114-37699
39	51	32.9	879	17	US-10-425-115-239018
40	51	32.9	903	15	US-10-282-122A-57086
41	50.5	32.6	45	14	US-10-043-487-243
42	50.5	32.6	209	17	US-10-739-930-10054
43	50.5	32.6	963	14	US-10-080-608A-20
44	50.5	32.6	963	14	US-10-080-608A-22
45	50.5	32.6	963	14	US-10-370-685-109

ALIGNMENTS

RESULT 1
US-10-092-750-24
; Sequence 24, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-XL
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-24

Query Match 100.0%; Score 155; DB 14; Length 33;
Best Local Similarity 100.0%; Pred. No. 3.5e-13;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GFLAAEQDIREIRKVVQSLQETAREVLTLLQ 33
Db 1 GFLAAEQDIREIRKVVQSLQETAREVLTLLQ 33

RESULT 2
US-09-925-297-499
; Sequence 499, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:

```

; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05989
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 499
; LENGTH: 253
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-297-499

```

```

Query Match          99.4%; Score 154; DB 9; Length 253;
Best Local Similarity 97.0%; Pred. No. 5,1e-12;
Matches 32; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GFLLAEQDIREIRKVVQSLEQTAREVLTLLQG 33
Db 37 GFLLAEQDIREIRKVVQSLEQTAREVLTLLQG 69

```

```

RESULT 3
US-10-108-260A-3640
; Sequence 3640, Application US/10108260A
; Publication No. US2004000560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. US2004000560A1el full length cDNA
; FILE REFERENCE: HL-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3640
; LENGTH: 429
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-108-260A-3640

```

```

Query Match          37.4%; Score 58; DB 15; Length 429;
Best Local Similarity 42.9%; Pred. No. 34;
Matches 12; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

```

```

QY 4 AAQDIREIRKVVQSLEQTAREVLTLL 31
Db 85 AAQDIREIRKVVQSLEQTAREVLTLL 112

```

```

RESULT 4
US-10-156-761-13249
; Sequence 13249, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02

```

```

; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13249
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13249

```

```

Query Match          37.4%; Score 58; DB 14; Length 513;
Best Local Similarity 39.3%; Pred. No. 42;
Matches 11; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

```

```

QY 4 AAQDIREIRKVVQSLEQTAREVLTLL 31
Db 339 AAQDIREIRKVVQSLEQTAREVLTLL 366

```

```

RESULT 5
US-10-437-963-106139
; Sequence 106139, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 106139
; LENGTH: 234
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_10610C.1.pap
US-10-437-963-106139

```

```

Query Match          36.8%; Score 57; DB 16; Length 234;
Best Local Similarity 44.4%; Pred. No. 23;
Matches 12; Conservative 7; Mismatches 8; Indels 0; Gaps 0;

```

```

QY 5 AAQDIREIRKVVQSLEQTAREVLTLL 31
Db 59 SEDALKEIRRVKEEEQSMREALGLV 85

```

```

RESULT 6
US-10-476-924-22
; Sequence 22, Application US/10476924
; Publication No. US20040152093A1
; GENERAL INFORMATION:
; APPLICANT: YUE, Henry; DING, Li;
; APPLICANT: BAUGHN, Mariah R.; LAL, Preeti G.;
; APPLICANT: YUE, Huibin; HAFALIA, April J.A.;
; APPLICANT: LEE, Ernestine A.; ISON, Craig H.;
; APPLICANT: BECHA, Shanya D.; GURURAJAN, Rajasopal;
; APPLICANT: EMERLING, Brooke M.; GRIFFIN, Jennifer A.;
; APPLICANT: TANG, Y. Tom; LU, Dyung Aina M.;
; APPLICANT: YAO, Monique G.; CHAWLA, Narinder K.;
; APPLICANT: RAMKUMAR, Jayalaxmi; GANDHI, Ameena R.;
; APPLICANT: LEE, Soo Yeun; RICHARDSON, Thomas W.;
; APPLICANT: YANG, Junming; ELLIOTT, Vicki S.;
; APPLICANT: LU, Yan; THANGRAVELU, Kavitha;
; APPLICANT: HE, Ann; AZIMZAI, Yalda;
; APPLICANT: RAUMANN, Brigitte E.; SWARNKAR, Anita;
; APPLICANT: BURFORD, Neil
; TITLE OF INVENTION: NUCLEIC ACID-ASSOCIATED PROTEINS

```



```
; FILE REFERENCE: PF-0960 USN
; CURRENT APPLICATION NUMBER: US/10/476,924
; CURRENT FILING DATE: 2003-11-04
; PRIOR APPLICATION NUMBER: PCT/US02/14276
; PRIOR FILING DATE: 2002-05-02
; PRIOR APPLICATION NUMBER: US 60/288,598
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: US 60/291,776
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: US 60/292,172
; PRIOR FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: US 60/293,564
; PRIOR FILING DATE: 2001-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PERL Program
; SEQ ID NO 22
; LENGTH: 751
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: 1532441CD1
US-10-476-924-22

Query Match 36.1%; Score 56; DB 16; Length 751;
Best Local Similarity 38.7%; Pred. No. 1.2e+02;
Matches 12; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

QY 3 LAQAQDIREIRKVVQSLQEQTAREVLTILQ 33
Db 267 LAQKQVIGLQRAHVEAEEAERRLAELEG 297

RESULT 7
US-10-177-293-391
; Sequence 391, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Glatt, Karen
; APPLICANT: Zhao, Xumel
; APPLICANT: Gannavarpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Bast Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Pusztai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul B.
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
```

```
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 391
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-391

Query Match 36.1%; Score 56; DB 14; Length 862;
Best Local Similarity 33.3%; Pred. No. 1.4e+02;
Matches 13; Conservative 11; Mismatches 9; Indels 6; Gaps 1;

QY 1 GFLAERQDIREIRKVVQSLQEQTAREVLTILQ 33
Db 616 GLSQAQRDVQEQMAVLMSREQVSEELVRLQKNDLSLQ 654

RESULT 8
US-10-408-765A-2447
; Sequence 2447, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2447
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-2447

Query Match 36.1%; Score 56; DB 16; Length 862;
Best Local Similarity 33.3%; Pred. No. 1.4e+02;
Matches 13; Conservative 11; Mismatches 9; Indels 6; Gaps 1;

QY 1 GFLAERQDIREIRKVVQSLQEQTAREVLTILQ 33
Db 616 GLSQAQRDVQEQMAVLMSREQVSEELVRLQKNDLSLQ 654

RESULT 9
US-10-411-066-40
; Sequence 40, Application US/10411066
; Publication No. US20040005678A1
; GENERAL INFORMATION:
; APPLICANT: KEASLING, JAY
; APPLICANT: MARTIN, VINCENT
; APPLICANT: PITREA, DOUGLAS
; APPLICANT: WITHERS III, SYDNOR T.
; APPLICANT: NEWMAN, JACK
; TITLE OF INVENTION: BIOSYNTHESIS OF AMORPHA-4,11-DIENE
; FILE REFERENCE: 2000-0007.20
; CURRENT APPLICATION NUMBER: US/10/411,066
; CURRENT FILING DATE: 2000-04-09
; PRIOR APPLICATION NUMBER: 10/006,909
; PRIOR FILING DATE: 2001-12-06
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 40
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Artemisia annua
```

US-10-411-066-40

Query Match 35.2%; Score 54.5; DB 15; Length 551;
Best Local Similarity 41.9%; Pred. No. 1.3e+02;
Matches 13; Conservative 8; Mismatches 9; Indels 1

Qy 2 FLAEPQDIREIRKWQSL-EQTAREVLTLL 31
|||:|||:|||:|||:|||:|||

Dp 24 FLAYDDEOEGVEOVIKDKSEVKSELLTAL 54

RESULT 10

```

US-10-425-115-236205
; Sequence 236205, Application US/10425115
; Publication NO. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalig, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 236205
; LENGTH: 217
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MFT4577 146C.1.pep

```

Query Match 34.8%; Score 54; DB 17; Length 217;
Best Local Similarity 46.2%; Pred. No. 51;
Matches 12; Conservative 5; Mismatches 9; Indels

QY 5 AEQDIREERKVVQVQSLEQTAREVLT 30
: | : | | | | : | | | | |
Db 59 SEDAMKEEIRRVKEEEEOAMREALGL 84

RESIST 11

```

US-10-425-115-357714
; Sequence 357714, Application US/10425115
; Publication NO. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 357714
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Cline ID: MRT4577_89404C.1.pep
US-10-425-115-357714

```

Query Match 34.8%; Score 54; DB 17; Length 221;
Best Local Similarity 48.0%; Pred. No. 52;
Matches 12: Conservative 4; Mismatches 9; Indels

QY 6 EQDIRREEIRKVQSLEQTAREVLT 30
| : | | | : | | | |
D5 60 EDALKEEIRRVKEEEEQAMREALGL 84

RESULT 12

US-10-425-114-60374
; Sequence 60374, Application US/10425114
; Publication No. US20040034888A1

```

: GENERAL INFORMATION:
: APPLICANT: Liu, Jingdong
: APPLICANT: Zhou, Yihua
: APPLICANT: Kovalic, David K.
: APPLICANT: Screen, Steven E
: APPLICANT: Tabaska, Jack E
: APPLICANT: Cao, Yongwei
: TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
: FILE REFERENCE: 38-21(53313)B
: CURRENT APPLICATION NUMBER: US/10/425,114
: CURRENT FILING DATE: 2003-04-28
: NUMBER OF SEQ ID NOS: 73128
: SEQ ID NO 60374
: LENGTH: 230
: TYPE: prt
: ORGANISM: Zea mays
: FEATURE:
: OTHER INFORMATION: Clone ID: LIB3596-046-C12_FLI.pep
: US-10-425-114-60374

```

```
Query Match      34.8%; Score 54; DB 15; Length 230;
Best Local Similarity 48.0%; Pred. NO. 55;
Matches 12; Conservative 4; Mismatches 9; Indels
```

QY 6 EQDIR EEIRKVVQSLEQTAREVLT L 30
 | : : : : : : : | | |
Dp 69 EDALKEEIRRVKEEEEQAMREALGL 93

RESULT 13

```

RES001.13
US-10-767-701-43357
; Sequence 43357, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 43357
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-C24743_1.pep
US-10-767-701-43357

```

Query Match 34.8%; Score 54; DB 16; Length 230;
Best Local Similarity 48.0%; Pred. No. 55;
Matches 12: Conservative 4; Mismatches 9; Indels

Qy	6	EQDIREERKVVQSLEQTAREVLT	30
		: : : : : : : : : : : :	
Dp	60	EDALKEEIRRVKKEEEQAMREALGL	84

RESULT 14

US-10-425-415-357715 ; Sequence 357715, Application US/104255115
; Publication No. US200421472A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.

```

; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 357715
; LENGTH: 230
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRP4577_89405C.1.pap
US-10-425-115-357715

```

```

Query Match      34.8%; Score 54; DB 17; Length 230;
Best Local Similarity 48.0%; Pred. No. 55;
Matches 12; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

```

```

QY      6  EODIREIRKVVQSLEOTAREVLT 30
      :|||||:|:|:|:|:|:|:|:|:|:|:|
Db      60  EDALKEIRRVKBEQEAMREALGL 84

```

```

RESULT 15
US-10-425-114-59221
; Sequence 59221, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 59221
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700208180_FLI.pap
US-10-425-114-59221

```

```

Query Match      34.8%; Score 54; DB 15; Length 239;
Best Local Similarity 48.0%; Pred. No. 57;
Matches 12; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

```

```

QY      6  EODIREIRKVVQSLEOTAREVLT 30
      :|||||:|:|:|:|:|:|:|:|:|:|:|
Db      69  EDALKEIRRVKBEQEAMREALGL 93

```

```

Search completed: November 11, 2004, 01:28:06
Job time : 36.5008 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:29:32 ; Search time 7.35233 Seconds
(without alignments)
431.857 Million cell updates/sec

Title: US-10-092-750-24
Perfect score: 155
Sequence: 1 GFLAAEQDIRREIRKVVQSLEQTAREVLTLLQG 33

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_79:*
1: Pirl:*
2: Pirl2:*
3: Pirl3:*
4: Pirl4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	154	99.4	228	2 S51738	translin - human
2	62	40.0	261	2 S72639	hypothetical prima
3	57	36.8	712	2 E83067	probable chemotaxi
4	56	36.1	714	2 A83314	probable chemotaxi
5	55	35.5	115	2 S01281	hypothetical prote
6	54	34.8	1383	2 T13052	guanine nucleotide
7	54	34.8	1920	2 A53188	pericentrin - mous
8	52	33.5	124	2 T10919	3C3.10 protein - S
9	52	33.5	182	2 B84387	hypothetical prote
10	52	33.5	264	2 F71466	hypothetical prote
11	52	33.5	390	1 A42724	transcription init
12	52	33.5	390	2 AG2463	transcription init
13	52	33.5	443	2 T47963	hypothetical prote
14	52	33.5	467	2 D72451	hypothetical prote
15	52	33.5	781	2 C69432	signal-transducing
16	51	32.9	133	2 H82366	conserved hypothet
17	51	32.9	646	2 AC3214	methyl-accepting c
18	51	32.9	664	2 A11459	methionyl-tRNA syn
19	51	32.9	790	2 F86936	probable integral
20	51	32.9	2954	2 T14156	kinesin-related pr
21	50.5	32.6	105	2 C75487	hypothetical prote
22	50.5	32.6	134	2 I47010	gastrin-releasing
23	50.5	32.6	220	1 NXEC	exclusion-determin
24	50.5	32.6	881	2 I84737	kinesin heavy chai
25	50.5	32.6	963	1 A41919	kinesin heavy chai
26	50	32.3	269	2 G71188	hypothetical prote
27	50	32.3	575	2 F70429	RNA polymerase sig
28	50	32.3	636	2 S76730	hypothetical prote
29	50	32.3	649	2 S16682	methionine-tRNA 11

30	50	32.3	723	2 T49985	oligopeptidase A-1
31	50	32.3	726	2 D72392	cation-transportin
32	50	32.3	753	2 AB2729	succinoglycan bios
33	50	32.3	753	2 C97510	hypothetical prote
34	50	32.3	760	2 C70956	probable ftsH prot
35	50	32.3	1170	2 A72287	hypothetical prote
36	50	32.3	1177	2 B75150	chromosome segrega
37	49.5	31.9	426	2 C72515	hypothetical prote
38	49.5	31.9	1137	2 T02097	glutamate synthase
39	49.5	31.9	1629	2 C84839	ferredoxin-depende
40	49	31.6	192	2 C97315	uncharacterized pr
41	49	31.6	218	2 G71197	hypothetical prote
42	49	31.6	233	2 F75213	hypothetical prote
43	49	31.6	269	2 G75148	hypothetical prote
44	49	31.6	291	2 A71194	hypothetical prote
45	49	31.6	320	2 S66176	ACC oxidase (clone

ALIGNMENTS

RESULT 1

S51738
translin - human
C;Species: Homo sapiens (man)
C;Date: 07-May-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C;Accession: S51738
R;Aoki, K.; Sugano, T.; Nakahara, K.; Kuge, O.; Omori, A.; Kasai, M.
submitted to the EMBL Data Library, April 1994
A;Description: Molecular basis of chromosome fragility.
A;Reference number: S51738
A;Accession: S51738
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-228 <AOK>
A;Cross-references: UNIPROT:Q15631; EMBL:X78627; NID:g607129; PIDN:CAA55341.1; PID:g60711
C;Genetics: GDB:TSN; TRSLN
A;Gene: GDB:TSN; TRSLN
A;Cross-references: GDB:731978; OMIM:600575
A;Map position: 10pter-10qter

Query Match 99.4%; Score 154; DB 2; Length 228;
Best Local Similarity 97.0%; Pred. No. 4.6e-11;
Matches 32; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GFLAAEQDIRREIRKVVQSLEQTAREVLTLLQG 33
|||||
DB 12 GFLAAEQDIRREIRKVVQSLEQTAREVLTLLQG 44

RESULT 2

S72639
hypothetical primary sigma factor - Thermoanaerobacterium thermosulfurigenes
A;Species: Thermoanaerobacterium thermosulfurigenes
C;Date: 29-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 09-Jul-2004
C;Accession: S72639
R;Matuschek, M.; Sahm, K.; Bahl, H.
submitted to the EMBL Data Library, March 1996
A;Description: Characterization of genes from Thermoanaerobacterium thermosulfurigenes
A;Reference number: S72639
A;Accession: S72639
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-261 <MAT>
A;Cross-references: UNIPROT:P77996; EMBL:U50951; NID:g1542972; PIDN:AAB08043.1; PID:g1541
C;Genetics:
A;Gene: sigma
C;Superfamily: Lyme disease spirochete probable RNA polymerase sigma factor; transcriptio
F;30-255/Domain: transcription initiation factor sigma katF homology <KTF>

Query Match 40.0%; Score 62; DB 2; Length 261;
Best Local Similarity 38.7%; Pred. No. 4.2;
Matches 12; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

QY 3 LAAEQDIREIRKVVQSLQEQTAREVLTLLQ 33
 A:Title: Sequence of figwort mosaic virus DNA (caulimovirus group).
 A:Reference number: S01279; MUID:88040466; PMID:3671088
 A:Accession: S01281
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-115 <RIC>
 A:Cross-references: UNIPROT:P09522; EMBL:X06166; NID:G58808; PIDN:CAA29525.1; PID:G58811
 C:Superfamily: cauliflower mosaic virus DNA-binding protein

Db 183 MAIEKEIKKIRKILDSUPERMILKVRFG 213

RESULT 3
 E83067
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
 C:Accession: E83067
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, A.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: E83067
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-712 <STO>
 A:Cross-references: UNIPROT:Q9HVF8; GB:AE004877; GB:AE004091; NID:G9950878; PIDN:AAG0802
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA4633

Query Match 36.8%; Score 57; DB 2; Length 712;
 Best Local Similarity 28.6%; Pred. No. 47;
 Matches 12; Conservative 10; Mismatches 10; Indels 10; Gaps 1;

QY 1 GFLAEQDIR-----BEIRKVVQSLQEQTAREVLTLLQ 32
 A:Title: Sequence of figwort mosaic virus DNA (caulimovirus group).
 A:Reference number: S01279; MUID:88040466; PMID:3671088
 A:Accession: S01281
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-115 <RIC>
 A:Cross-references: UNIPROT:P09522; EMBL:X06166; NID:G58808; PIDN:CAA29525.1; PID:G58811
 C:Superfamily: cauliflower mosaic virus DNA-binding protein

Db 576 GFAVVADEVRLAQTQQTAEIQSMIQLOQGRDVRVKNQ 617

RESULT 4
 A83314
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
 C:Accession: A83314
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Berman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, A.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: A83314
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-714 <STO>
 A:Cross-references: UNIPROT:Q9HVF8; GB:AE004877; GB:AE004091; NID:G9950878; PIDN:AAG0802
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA2854

Query Match 36.1%; Score 56; DB 2; Length 714;
 Best Local Similarity 40.7%; Pred. No. 62;
 Matches 11; Conservative 7; Mismatches 9; Indels 9; Gaps 0;

QY 7 QDIREIRKVVQSLQEQTAREVLTLLQ 33
 A:Title: Sequence of figwort mosaic virus DNA (caulimovirus group).
 A:Reference number: S01279; MUID:88040466; PMID:3671088
 A:Accession: S01281
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-115 <RIC>
 A:Cross-references: UNIPROT:P09522; EMBL:X06166; NID:G58808; PIDN:CAA29525.1; PID:G58811
 C:Superfamily: cauliflower mosaic virus DNA-binding protein

Db 594 QDSVEQIRVIGLQGRDVRVKNQ 620

RESULT 5
 S01281
 C:Species: figwort mosaic virus
 C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
 C:Accession: S01281
 R:Richins, R.D.; Scholthof, H.B.; Shepherd, R.J.

Query Match 34.8%; Score 54; DB 2; Length 1920;
 Best Local Similarity 46.2%; Pred. No. 38+02;
 Matches 12; Conservative 6; Mismatches 8; Indels 8; Gaps 0;

QY 5 AEQDIREIRKVVQSLQEQTAREVLTLLQ 30
 A:Title: Sequence of figwort mosaic virus DNA (caulimovirus group).
 A:Reference number: S01279; MUID:88040466; PMID:3671088
 A:Accession: S01281
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-115 <RIC>
 A:Cross-references: UNIPROT:P09522; EMBL:X06166; NID:G58808; PIDN:CAA29525.1; PID:G58811
 C:Superfamily: cauliflower mosaic virus DNA-binding protein

Db 1132 SEQQLILEDLRKQLEQAARELLTL 1157

Nucleic Acids Res. 15, 8451-8466, 1987
 A:Title: Sequence of figwort mosaic virus DNA (caulimovirus group).
 A:Reference number: S01279; MUID:88040466; PMID:3671088
 A:Accession: S01281
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-115 <RIC>
 A:Cross-references: UNIPROT:P09522; EMBL:X06166; NID:G58808; PIDN:CAA29525.1; PID:G58811
 C:Superfamily: cauliflower mosaic virus DNA-binding protein

Query Match 35.5%; Score 55; DB 2; Length 115;
 Best Local Similarity 47.6%; Pred. No. 12;
 Matches 10; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 12 EIRKVVQSLQEQTAREVLTLLQ 32
 A:Title: Sequence of figwort mosaic virus DNA (caulimovirus group).
 A:Reference number: S01279; MUID:88040466; PMID:3671088
 A:Accession: S01281
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-115 <RIC>
 A:Cross-references: UNIPROT:P09522; EMBL:X06166; NID:G58808; PIDN:CAA29525.1; PID:G58811
 C:Superfamily: cauliflower mosaic virus DNA-binding protein

Db 7 EIRELIQSLTKTANEIKAMLE 27

RESULT 6
 T13052
 C:Species: Drosophila melanogaster
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Mar-2004
 C:Accession: T13052
 R:Shimizu, S.; Gaul, U.; Hanafusa, H.
 submitted to the EMBL Data Library, March 1998
 A:Description: cDNA sequence of Drosophila melanogaster DC3G gene.
 A:Reference number: Z17593
 A:Accession: T13052
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1383 <LSH>
 A:Cross-references: EMBL:AF053358; NID:G3582764; PID:G3582765; PIDN:AAC35280.1
 C:Genetics:
 A:Gene: DC3G
 A:Cross-references: FlyBase:FBgn0026145
 A:Map position: X
 F:1147-1376/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 34.8%; Score 54; DB 2; Length 1383;
 Best Local Similarity 43.5%; Pred. No. 2.2e+02;
 Matches 10; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

QY 11 EIRKVVQSLQEQTAREVLTLLQ 33
 A:Title: Sequence of figwort mosaic virus DNA (caulimovirus group).
 A:Reference number: S01279; MUID:88040466; PMID:3671088
 A:Accession: S01281
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-115 <RIC>
 A:Cross-references: UNIPROT:P09522; EMBL:X06166; NID:G58808; PIDN:CAA29525.1; PID:G58811
 C:Superfamily: cauliflower mosaic virus DNA-binding protein

Db 279 ENVREVIDLEDAVNLVTLAQ 301

RESULT 7
 A53188
 C:Species: Mus musculus (house mouse)
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 09-Jul-2004
 C:Accession: A53188
 R:Doxsey, S.J.; Stein, P.; Evans, L.; Calarco, P.D.; Kirschner, M.
 Cell 76, 639-650, 1994
 A:Title: Pericentrin, a highly conserved centrosome protein involved in microtubule organization
 A:Reference number: A53188; MUID:94170365; PMID:8124707
 A:Accession: A53188
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1920 <DOX>
 A:Cross-references: UNIPROT:P48725; GB:U05623; NID:G458667; PIDN:AAA17886.1; PID:G458668
 C:Keywords: coiled coil

Query Match 34.8%; Score 54; DB 2; Length 1920;
 Best Local Similarity 46.2%; Pred. No. 38+02;
 Matches 12; Conservative 6; Mismatches 8; Indels 8; Gaps 0;

QY 5 AEQDIREIRKVVQSLQEQTAREVLTLLQ 30
 A:Title: Sequence of figwort mosaic virus DNA (caulimovirus group).
 A:Reference number: S01279; MUID:88040466; PMID:3671088
 A:Accession: S01281
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-115 <RIC>
 A:Cross-references: UNIPROT:P09522; EMBL:X06166; NID:G58808; PIDN:CAA29525.1; PID:G58811
 C:Superfamily: cauliflower mosaic virus DNA-binding protein

Db 1132 SEQQLILEDLRKQLEQAARELLTL 1157

A;Cross-references: UNIPROT:O84819; GB:AE001354; GB:AE001273; NID:g3329280; PIDN:AAC6849
A;Experimental source: serotype D, strain UW-3/Cx
C;Genetics:
A;Gene: CT813

Query Match 33.5%; Score 52; DB 2; Length 264;
Best Local Similarity 34.8%; Pred. No. 65;
Matches 8; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

QY 6 EQDIEEIRKVVQSLEQTAREVL 28
:|::|||:|::|||:|::|||:
Db 225 KQELREELDKAIENLDEMAYEAM 247

RESULT 11
A42724
transcription initiation factor sigma sigA - Anabaena sp. (strain PCC 7120)
C;Species: Anabaena sp.
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C;Accession: A42724
J.Brahamsba, B.; Haselkorn, R.
R. Bacteriol. 173, 2442-2450, 1991
A;Title: Isolation and characterization of the gene encoding the principal sigma factor
A;Reference number: A42724; MUID:91193199; PMID:1901566
A;Accession: A42724
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-390

A;Cross-references: UNIPROT:P26683; GB:M60046; NID:g142107; PIDN:AAA22043.1; PID:g142108
C;Superfamily: transcription initiation factor sigma 43; transcription initiation factor
C;Keywords: DNA binding; sigma factor; transcription initiation
F.160-385/Domain: transcription initiation factor sigma katF homology <XTF>

Query Match 33.5%; Score 52; DB 1; Length 390;
Best Local Similarity 44.0%; Pred. No. 98;
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 9 IREEIRKVQSLEQTAREVLTLLQG 33
:|::|||:|::|||:|::|||:
Db 318 LREDLEKVLDSLSPRRDVLRVYG 342

RESULT 12
AG2463
transcription initiation factor sigma sigA [imported] - Nostoc sp. (strain PCC 7120)
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AG2463
R.Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AG2463
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-390 <KUR>
A;Cross-references: UNIPROT:P26683; GB:BA000019; PIDN:BAB76962.1; PID:g17134402; GSPDB:G
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: sigA
C;Superfamily: transcription initiation factor sigma 43; transcription initiation factor
C;Keywords: transcription initiation

Query Match 33.5%; Score 52; DB 2; Length 390;
Best Local Similarity 44.0%; Pred. No. 98;
Matches 11; Conservative 5; Mismatches 9; Indels 0; Gaps 0;

QY 9 IREEIRKVQSLEQTAREVLTLLQG 33
:|::|||:|::|||:|::|||:
Db 318 LREDLEKVLDSLSPRRDVLRVYG 342

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 40.2953 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-24
Perfect score: 155
Sequence: 1 GFLAAEQDIREIRKVVQSLAQAREVLTLLQG 33

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_02.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	154	99.4	228	1 TSN_CRIGR	P97891 cricetus
2	154	99.4	228	1 TSN_HUMAN	Q15631 homo sapien
3	154	99.4	228	1 TSN_MOUSE	Q62348 mus musculus
4	154	99.4	228	1 TSN_MOUSE	Q715Y3 rattus norv
5	154	99.4	228	2 AAF60295	Aaf60295 mus muscu
6	154	99.4	228	2 AAF91387	Aaf91387 rattus no
7	154	99.4	228	2 BAB27152	Bab27152 mus muscu
8	129	83.2	154	2 Q7ZXK0	Q7zxk0 xenopus lae
9	129	83.2	228	2 Q1AM5	Q1am5 xenopus lae
10	127	81.9	229	1 TSN_CHICK	P79769 gallus gall
11	101.5	65.5	278	2 Q7TP02	Q7tp02 rattus norv
12	71	45.8	287	2 Q7Q729	Q7q729 anopheles g
13	62	40.0	261	2 P77996	P77996 thermoaer
14	60.5	39.0	139	2 Q7PK34	Q7pk34 anopheles g
15	60	38.7	325	2 Q6DF92	Q6df92 xenopus lae
16	59	38.1	235	2 Q9V5M0	Q9v5m0 drosophila
17	59	38.1	391	2 Q8PXN8	Q8pxn8 methanosarc
18	59	38.1	964	2 Q8S228	Q8s228 setaria ita
19	58.5	37.7	480	2 Q815U7	Q815u7 bacillus ce
20	58	37.4	149	2 Q74EJ4	Q74eu4 geobacter s
21	58	37.4	149	2 AAR34195	Aar34195 geobacter
22	58	37.4	418	2 Q9F413	Q9f413 mycoplasma
23	58	37.4	429	2 Q7L5L2	Q7l5l2 homo sapien
24	58	37.4	429	2 Q8NFZ5	Q8nfz5 homo sapien
25	58	37.4	429	2 Q9BQR6	Q9bqt6 homo sapien
26	58	37.4	429	2 AAH02740	Aah02740 homo sapi
27	58	37.4	513	2 Q82B16	Q82bi6 streptomyce
28	57	36.8	234	2 Q8YZ17	Q8yz17 oryza sativ
29	57	36.8	234	2 BAD13130	Bad13130 oryza sat
30	57	36.8	628	2 Q87FM8	Q87fm8 vibrio para
31	57	36.8	712	2 Q9HV78	Q9hvf8 pseudomonas

ALIGNMENTS

RESULT 1

ID	TSN_CRIGR	STANDARD	PRT	228 AA
AC	P97891			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUL-1999 (Rel. 38, Last sequence update)			
DT	05-JUL-2004 (Rel. 44, Last annotation update)			
DE	Translin.			
GN	Name-TSN:			
OS	Cricetulus griseus (Chinese hamster).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;			
OC	Cricetulus.			
OX	NCBI_TaxID=10029;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RT	TISSUE=Ovary;			
RA	Kasai M.;			
RL	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: DNA-binding protein that specifically recognizes consensus sequences at the breakpoint junctions in chromosomal translocations, mostly involving immunoglobulin (Ig)/T-cell receptor gene segments. Seems to recognize single-stranded DNA ends generated by staggered breaks occurring at recombination hot spots (by similarity).			
CC	-!- SUBUNIT: Forms a multimeric ring-shaped structure (By similarity).			
CC	-!- SUBCELLULAR LOCATION: Nuclear.			
CC	-!- SIMILARITY: Belongs to the translin family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X88066; CAA66693.1; -.			
DR	HSSP; Q62348; IKEY.			
DR	InterPro; IPR002848; Translin.			
DR	Pfam; PF01997; Translin; 1.			
KW	DNA-binding; Nuclear protein.			
FT	DOMAIN 177 198 Leucine-zipper (Potential).			
SQ	SEQUENCE 228 AA; 26171 MW; 29AAAFDB45D73F69C CRC64;			
Query Match	99.4%; Score 154; DB 1; Length 228;			
Best Local Similarity	97.0%; Pred. No. 1.8e-09;			
Matches	32; Conservative 1; Mismatches 0; Indels 0; Gaps 0;			

QY 1 GFLAAEQDIREIRKVVQSLAQAREVLTLLQG 33

DB 12 GFLAAEQDIREIRKVVQSLAQAREVLTLLQG 44

RESULT 2
 ID TSN HUMAN STANDARD; PRT: 228 AA.
 AC Q15631;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Translin.
 GN Name=TSN;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95392568; PubMed=7663511;
 RA Aoki K., Suzuki K., Sugano T., Tasaka T., Nakahara K., Kuge O.,
 RA Omori A., Kasai M.;
 RT "A novel gene, Translin, encodes a recombination hotspot binding
 RT protein associated with chromosomal translocations.";
 RL Nat. Genet. 10:167-174(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin D.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S.S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
 RA Rodriguez A.C., Goughwood J.J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 23-215 FROM N.A.
 RX MEDLINE=97386594; PubMed=9244443;
 RA Aoki K., Inazawa J., Takahashi T., Nakahara K., Kasai M.;
 RT "Genomic structure and chromosomal localization of the gene encoding
 RT translin, a recombination hotspot binding protein.";
 RL Genomics 43:237-241(1997).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=97269049; PubMed=9111049;
 RA Kasai M., Matsuzaki T., Katayangi K., Omori A., Maziarz R.T.,
 RA Strominger J.L., Aoki K., Suzuki K.;
 RT "The translin ring specifically recognizes DNA ends at recombination
 RT hot spots in the human genome.";
 RL J. Biol. Chem. 272:11402-11407(1997).
 CC -!- FUNCTION: DNA-binding protein that specifically recognizes
 CC consensus sequences at the breakpoint junctions in chromosomal
 CC translocations, mostly involving immunoglobulin (Ig)/T-cell
 CC receptor gene segments. Seems to recognize single-stranded DNA
 CC ends generated by staggered breaks occurring at recombination hot
 CC spots.
 CC -!- SUBUNIT: Forms a multimeric ring-shaped structure.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: Belongs to the translin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X78627; CAA55341.1; -;
 DR EMBL: BC002359; AB02359.1; -;
 DR EMBL: Y12563; CAA73150.1; -;
 DR EMBL: Y12564; CAA73150.1; JOINED.
 DR EMBL: Y12565; CAA73150.1; JOINED.
 DR EMBL: Y12566; CAA73150.1; JOINED.
 DR EMBL: Y12567; CAA73150.1; JOINED.
 DR PIR: S51738; S51738.
 DR PDB: 1JJJ; X-ray; A/B/C/D=1-228.
 DR Genew; HGNC:12379; TSN.
 DR MIM: 600575; -;
 DR GO: GO:0005634; C:nucleus; TAS.
 DR GO: GO:0003677; F:DNA binding; TAS.
 DR GO: GO:0006310; P:DNA recombination; TAS.
 DR InterPro: IPR002848; Translin.
 DR Pfam: PF01997; Translin; 1.
 DR 3D-structure; DNA-binding; Nuclear protein.
 FT DOMAIN 177 198 Leucine-zipper (Potential).
 SQ SEQUENCE 228 AA; 26183 MW; 3CAAF20BED7C4939 CRC64;
 Query Match 99.4%; Score 154; DB 1; Length 228;
 Best Local Similarity 97.0%; Pred. No. 1.8e-09;
 Matches 32; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 GFLAARDREIRKVVQSLQETAREVLTLLQ 33
 Db 12 GFLAARDREIRKVVQSLQETAREVLTLLQ 44
 RESULT 3
 TSN_MOUSE
 ID TSN_MOUSE STANDARD; PRT: 228 AA.
 AC Q62348;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-OCT-2004 (Rel. 45, Last annotation update)
 DE Translin.
 GN Name=TSN;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Thymus;
 RX MEDLINE=97386594; PubMed=9244443;
 RA Aoki K., Inazawa J., Takahashi T., Nakahara K., Kasai M.;
 RT "Genomic structure and chromosomal localization of the gene encoding
 RT translin, a recombination hotspot binding protein.";
 RL Genomics 43:237-241(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FVB/N; TISSUE=Mammary gland;
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Ustin D.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S.S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
 RA Rodriguez A.C., Goughwood J.J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP SEQUENCE OF 23-215 FROM N.A.
 RX MEDLINE=97386594; PubMed=9244443;
 RA Aoki K., Inazawa J., Takahashi T., Nakahara K., Kasai M.;
 RT "Genomic structure and chromosomal localization of the gene encoding
 RT translin, a recombination hotspot binding protein.";
 RL Genomics 43:237-241(1997).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=97269049; PubMed=9111049;
 RA Kasai M., Matsuzaki T., Katayangi K., Omori A., Maziarz R.T.,
 RA Strominger J.L., Aoki K., Suzuki K.;
 RT "The translin ring specifically recognizes DNA ends at recombination
 RT hot spots in the human genome.";
 RL J. Biol. Chem. 272:11402-11407(1997).
 CC -!- FUNCTION: DNA-binding protein that specifically recognizes
 CC consensus sequences at the breakpoint junctions in chromosomal
 CC translocations, mostly involving immunoglobulin (Ig)/T-cell
 CC receptor gene segments. Seems to recognize single-stranded DNA
 CC ends generated by staggered breaks occurring at recombination hot
 CC spots.
 CC -!- SUBUNIT: Forms a multimeric ring-shaped structure.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: Belongs to the translin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzyzinski M.I., Skalska U., Smallos D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -!- FUNCTION: DNA-binding protein that specifically recognizes
 CC consensus sequences at the breakpoint junctions in chromosomal
 CC translocations, mostly involving immunoglobulin (Ig)/T-cell
 CC receptor gene segments. Seems to recognize single-stranded DNA
 CC ends generated by staggered breaks occurring at recombination hot
 CC spots.
 CC -!- SUBUNIT: Forms a multimeric ring-shaped structure.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: Belongs to the translin family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; X81464; CAA57222.1; -;
 CC EMBL; Y12568; CAA73151.1; -;
 CC EMBL; Y12569; CAA73151.1; JOINED.
 CC EMBL; Y12570; CAA73151.1; JOINED.
 CC EMBL; Y12571; CAA73151.1; JOINED.
 CC EMBL; Y12572; CAA73151.1; JOINED.
 CC EMBL; BC004615; AAH04615.1; -;
 CC PDB; 1KEY; X-ray; A/B/C/D=1-228.
 CC MGI; MGI:109263; Tsn.
 CC GO; GO:0005737; C:cytoplasm; IDA.
 CC GO; GO:0005634; C:nucleus; IDA.
 CC GO; GO:000515; F:protein binding; IPI.
 CC GO; GO:0003723; F:RNA binding; IDA.
 CC InterPro; IPR002848; Translin.
 CC Pfam; PF01997; Translin; 1.
 CC 3D-structure; DNA-binding; Nuclear protein.
 CC Leucine-zipper (Potential).
 FT DOMAIN 177 198
 FT HELIX 3 43
 FT TURN 44 45
 FT HELIX 55 75
 FT TURN 76 77
 FT TURN 85 87
 FT HELIX 88 90
 FT TURN 91 91
 FT HELIX 92 110
 FT HELIX 116 122
 FT TURN 123 124
 FT HELIX 137 161
 FT TURN 162 162
 FT HELIX 166 182
 FT TURN 183 183
 FT HELIX 189 195
 FT TURN 196 197
 FT HELIX 198 215
 FT TURN 216 216
 SQ SEQUENCE 228 AA; 26201 MW; 39PBB4FCC5AA375C CRC64;
 Query Match 99.4%; Score 154; DB 1; Length 228;
 Best Local Similarity 97.0%; Pred. No. 1.8e-09;
 Matches 32; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GFLAAEQDIREIRKVVQSLQCTAREVLTLLQ 33
 DB 12 GFLAAEQDIREIRKVVQSLQCTAREVLTLLQ 44
 RESULT 4
 ID Q71SV3 PRELIMINARY; PRT; 228 AA.
 Q71SV3

AC Q71SV3;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Translin.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Brain;
 RX MEDLINE=98344761; PubMed=98681436;
 RA Taira E., Finkenshtadt P.M., Baraban J.M.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF262356; AAF91387.1; -;
 DR InterPro; IPR002848; Translin.
 DR Pfam; PF01997; Translin; 1.
 SQ SEQUENCE 228 AA; 26171 MW; 29AAFD45D73F69C CRC64;
 Query Match 99.4%; Score 154; DB 2; Length 228;
 Best Local Similarity 97.0%; Pred. No. 1.8e-09;
 Matches 32; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GFLAAEQDIREIRKVVQSLQCTAREVLTLLQ 33
 DB 12 GFLAAEQDIREIRKVVQSLQCTAREVLTLLQ 44
 RESULT 5
 ID AAF60295 PRELIMINARY; PRT; 228 AA.
 AC AAF60295;
 DT 02-MAR-2004 (TrEMBLrel. 27, Created)
 DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
 DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
 DE RNA-binding protein.
 GN TBRBP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chennathukuzhi V.M., Lefrancois S., Morales C.R., Syed V., Hecht N.B.;
 RT "Elevated levels of the polyadenylation factor CstF 64 regulate the
 RT alternative processing of Testis Brain RNA-binding protein (TB-RBP)
 RT pre-mRNA in male germ cells";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF234179; AAF60295.1; -;
 SQ SEQUENCE 228 AA; 26201 MW; 39PBB4FCC5AA375C CRC64;
 Query Match 99.4%; Score 154; DB 2; Length 228;
 Best Local Similarity 97.0%; Pred. No. 1.8e-09;
 Matches 32; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GFLAAEQDIREIRKVVQSLQCTAREVLTLLQ 33
 DB 12 GFLAAEQDIREIRKVVQSLQCTAREVLTLLQ 44
 RESULT 6
 ID AAF91387 PRELIMINARY; PRT; 228 AA.
 AC AAF91387;
 DT 02-MAR-2004 (TrEMBLrel. 27, Created)

DT 02-MAR-2004 (TReMBLrel. 27, Last sequence update)
 DT 02-MAR-2004 (TReMBLrel. 27, Last annotation update)
 DE Translin.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=brain;
 RX MEDLINE=98344761; PubMed=9681436;
 RA Taira E., Finkentadt P.M., Baraban J.M.;
 RT "Identification of translin and trax as components of the G31 strand-
 specific DNA binding complex enriched in brain.";
 RL J. Neurochem. 71:471-477(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=brain;
 RA Taira E., Finkentadt P.M., Baraban J.M.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF262356; AAF91387.1;
 SQ SEQUENCE 228 AA; 26171 MW; 29AAPDB45D73F69C CRC64;

 Query Match 99.4%; Score 154; DB 2; Length 228;
 Best Local Similarity 97.0%; Pred. No. 1.8e-09;
 Matches 32; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 QY 1 GFLAEQDIREIRKVKVQSLEQTAREVLTLLQ 33
 |||||
 DB 12 GFLAEQDIREIRKVKVQSLEQTAREVLTLLQ 44
 |||||

 RESULT 7
 BAB27152
 ID BAB27152 PRELIMINARY; PRT; 228 AA.
 AC BAB27152;
 DT 14-APR-2004 (TReMBLrel. 27, Created)
 DT 14-APR-2004 (TReMBLrel. 27, Last sequence update)
 DE ES cells cDNA, RIKEN full-length enriched library, clone:2410083E08
 DE product:translin, full insert sequence.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium,
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=21085660; PubMed=11217851;
 RA RIKEN FANTOM Consortium,
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=99279253; PubMed=10349636;
 RA Carninci P., Hayashizaki Y.;
 RT "High-efficiency full-length cDNA cloning.";
 RL Meth. Enzymol. 303:19-44(1999).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=20499374; PubMed=11042159;
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,

RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes.";
 RL Genome Res. 10:1617-1630(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
 RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
 RA Fujiwaka S., Inoue K., Togawa Y., Iizawa M., Ohara E., Watahiki M.,
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
 RT "RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer.";
 RL Genome Res. 10:1757-1771(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J;
 RX MEDLINE=20530913; PubMed=11076861;
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
 RA Inotani K., Ishii Y., Itoh M., Iizawa M., Kasukawa T., Kato H.,
 RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai M.,
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
 RA Muramatsu M., Hayashizaki Y.;
 RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK010737; BAB27152.1;
 SQ SEQUENCE 228 AA; 26201 MW; 39FBB4FCC5AA375C CRC64;

 Query Match 99.4%; Score 154; DB 2; Length 228;
 Best Local Similarity 97.0%; Pred. No. 1.8e-09;
 Matches 32; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

 QY 1 GFLAEQDIREIRKVKVQSLEQTAREVLTLLQ 33
 |||||
 DB 12 GFLAEQDIREIRKVKVQSLEQTAREVLTLLQ 44
 |||||

 RESULT 8
 Q7ZXK0
 ID Q7ZXK0 PRELIMINARY; PRT; 154 AA.
 AC Q7ZXK0;
 DT 01-JUN-2003 (TReMBLrel. 24, Created)
 DT 01-JUN-2003 (TReMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Tsn-prov protein.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Abramson R.D., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

```

P79769;
15-JUL-1999 (Rel. 38, Created)
15-JUL-1999 (Rel. 38, Last sequence update)
05-JUL-2004 (Rel. 44, Last annotation update)
Translin.
Name=TSN;
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
NCBI_TaxID=9031;
[1] NCBI_TaxID=9031;
RN NCBI_TaxID=9031;
RP SEQUENCE FROM N.A.
RC TISSUE=Thymus;
RA Kasei M.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
RL FUNCTION: DNA-binding protein that specifically recognizes
CC consensus sequences at the breakpoint junctions in chromosomal
CC translocations, mostly involving immunoglobulin (Ig)/T-cell
CC receptor gene segments. Seems to recognize single-stranded DNA
CC ends generated by staggered breaks occurring at recombination hot
CC spots (By similarity).
CC
CC -! SUBUNIT: Forms a multimeric ring-shaped structure (By similarity).
CC
CC -! SUBCELLULAR LOCATION: Nuclear.
CC
CC -! SIMILARITY: Belongs to the Translin family.
CC
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
-----
CC EMBL; X95074; CAA64470.1; -.
CC HSP; Q62348; 1KEY.
CC InterPro; IPR002848; Translin.
CC Pfam; PF01997; Translin; 1.
CC DNA-binding; Nuclear protein.
CC DOMAIN 177 198 Leucine-zipper (Potential).
CC SEQUENCE 229 AA; 25960 MW; E1DB50693A84D2C9 CRC64;
CC
Query Match 81.9%; Score 127; DB 1; Length 229;
Best Local Similarity 81.8%; Pred. No. 28-06;
Matches 27; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
QY 1 GFLAEQDIREERKVVQSLQETAREVLTLLQG 33
DB 12 GALTADQDIREERKVVQSLQETAREVLTLLQG 44
-----
RESULT 11
Q7TP02
ID Q7TP02 PRELIMINARY; PRT; 278 AA.
AC Q7TP02;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Da2-35.
DE OS Rattus norvegicus (Rat).
DE OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
DE OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
DE OC NCBI_TaxID=10116;
DE [1]
RN SEQUENCE FROM N.A.
RP Xu C.S., Li W.Q., Li Y.C., Wang L., Wang S.F., Han H.P., Wang G.P.,
RA Chai L.Q., Yuan J.Y., Yang K.J., Yan H.M., Chang C.F., Zhao L.F.,
RA Ma H., Shi J.B., Rahman S., Wang Q.N., Zhang J.B.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY325258; AAP92659.1; -.
DR InterPro; IPR002848; Translin.
DR Pfam; PF01997; Translin; 1.
SQ SEQUENCE 278 AA; 31390 MW; 3C429A619149B829 CRC64;

```

Query Match 65.5%; Score 101.5; DB 2; Length 278;
 Best Local Similarity 71.4%; Pred. No. 0.0018;
 Matches 25; Conservative 1; Mismatches 4; Indels 5; Gaps 1;

QY 4 AAEQDIREIRKVVQSLQEQAREVLTLLQ 33
 DB 57 AARQASRRARAPEIRKVVQSLQEQAREVLTLLQ 91

RESULT 12
 Q7Q729 PRELIMINARY; PRT; 287 AA.
 AC Q7Q729;
 DT 01-MAR-2004 (TReMBLrel. 26, Created)
 DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 DE AGCP6534 (Fragment).
 GN Name=agCG5365; ORFNames=ENSANGG00000019317;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
 OX NCBI_TaxID=180454;
 RN [1] SEQUENCE FROM N.A.
 RP STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAB01008960; EAA11822.1; -.
 DR InterPro; IPR002848; Translin.
 DR Pfam; PF01997; Translin; 1.
 FT NON_TER 1
 SQ SEQUENCE 287 AA; 32023 MW; D0DAD0F56C5B9037 CRC64;

Query Match 45.8%; Score 71; DB 2; Length 287;
 Best Local Similarity 41.9%; Pred. No. 5.2;
 Matches 13; Conservative 8; Mismatches 10; Indels 0; Gaps 0;

QY 2 FLAAEQDIREIRKVVQSLQEQAREVLTLLQ 32
 DB 50 YLVKQEQELTEIRDIVRIDQAAKEAIALQ 80

RESULT 13
 P77996 PRELIMINARY; PRT; 261 AA.
 AC P77996;
 DT 01-FEB-1997 (TReMBLrel. 02, Created)
 DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 DE Sigma factor.
 GN Name=sigA;
 OS Thermotoga bacter thermosulfurogenes (Clostridium
 OS thermosulfurogenes).
 OC Bacteria; Firmicutes; Clostridia; Thermotoga bacteriales;
 OC Thermotoga bacteriales; Thermotoga bacteriales;
 OX NCBI_TaxID=33950;
 RN [1] SEQUENCE FROM N.A.
 RP STRAIN=EM1;
 RA Matuschek M.; Sahn K.; Bahl H.;
 RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Sigma factors are initiation factors that promote the
 CC attachment of RNA polymerase to specific initiation sites and are
 CC then released (By similarity).
 CC -!- SIMILARITY: Belongs to the sigma-70 factor family.
 DR EMBL; U50951; AAB08043.1; -.
 DR PIR; S72639; S72639.
 DR HSSP; Q9EZ38; 1KU3.
 DR GO; GO:0003899; F:DNA-directed RNA polymerase activity; IEA.

DR GO; GO:0016987; F:sigma factor activity; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0016740; F:transferase activity; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR GO; GO:0006352; P:transcription initiation; IEA.
 DR InterPro; IPR009043; RNA_pol_sigma.
 DR InterPro; IPR007627; Sigma70 r3.
 DR InterPro; IPR007624; Sigma70 r3.
 DR InterPro; IPR007630; Sigma70 r4.
 DR InterPro; IPR000943; Sigma70.
 DR Pfam; PF04542; Sigma70_r2_1.
 DR Pfam; PF04539; Sigma70_r3_1.
 DR Pfam; PF04545; Sigma70_r4_1.
 DR PRINTS; PR00046; SIGMA70FCT.
 DR PROSITE; PS00715; SIGMA70 1.
 DR PROSITE; PS00716; SIGMA70 2; 1.
 KW DNA-binding; DNA-directed RNA polymerase; Sigma factor;
 KW Transcription regulation; Transferase
 SQ SEQUENCE 261 AA; 30627 MW; A99F6038F2842242 CRC64;

Query Match 40.0%; Score 62; DB 2; Length 261;
 Best Local Similarity 38.7%; Pred. No. 49;
 Matches 12; Conservative 10; Mismatches 9; Indels 0; Gaps 0;

QY 3 LAAEQDIREIRKVVQSLQEQAREVLTLLQ 33
 DB 183 MAIEKIKRILKILSLPERMILKMRFG 213

RESULT 14
 Q7PKA4 PRELIMINARY; PRT; 139 AA.
 ID Q7PKA4;
 AC Q7PKA4;
 DT 01-MAR-2004 (TReMBLrel. 26, Created)
 DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 DE ENSANGP00000023590 (Fragment).
 GN Name=ENSANGG00000020385;
 OS Anopheles gambiae str. PEST.
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.
 OX NCBI_TaxID=180454;
 RN [1] SEQUENCE FROM N.A.
 RP STRAIN=PEST;
 RA Anopheles Genome Sequencing Consortium;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 DR EMBL; AAB01008982; EAA43447.1; -.
 FT NON_TER 1
 FT NON_TER 139
 SQ SEQUENCE 139 AA; 16268 MW; 8DED247F114D8D00 CRC64;

Query Match 39.0%; Score 60.5; DB 2; Length 139;
 Best Local Similarity 53.6%; Pred. No. 40;
 Matches 15; Conservative 5; Mismatches 7; Indels 1; Gaps 1;

QY 5 ASQDIREIRKVVQSLQEQAREVLTLL 31
 DB 104 AENDRENGKVQVPERKAKDSTTM 131

RESULT 15
 Q6DF92 PRELIMINARY; PRT; 325 AA.
 ID Q6DF92;
 AC Q6DF92;
 DT 01-OCT-2004 (TReMBLrel. 28, Created)
 DT 01-OCT-2004 (TReMBLrel. 28, Last sequence update)
 DT 01-OCT-2004 (TReMBLrel. 28, Last annotation update)
 DE Hypothetical protein.
 OS Xenopus laevis (African clawed frog).

```
Query Match      38.7%; Score 60; DB 2; Length 325;
Best Local Similarity 44.8%; Pred. No. 1e+02;
Matches 13; Conservative 7; Mismatches 9; Indels 0; Gaps 0;
```

Search completed: November 10, 2004, 13:38:19
Job time : 42.2953 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:32:37 ; Search time 11.399 Seconds
(without alignments)
191.991 Million cell updates/sec

Title: US-10-092-750-25
Perfect score: 168
Sequence: 1 LDPVKDVLILSALRMLWAADFLDLPEQIG 33

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/1/iaa/5A.COMB.pap.*
2: /cgn2_6/ptodata/1/iaa/5B.COMB.pap.*
3: /cgn2_6/ptodata/1/iaa/6A.COMB.pap.*
4: /cgn2_6/ptodata/1/iaa/6B.COMB.pap.*
5: /cgn2_6/ptodata/1/iaa/PCTUS.COMB.pap.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	52	31.0	510	4	US-09-270-767-42036
2	50	29.8	491	4	US-09-248-796A-14421
3	49.5	29.5	306	4	US-09-134-000C-6601
4	49.5	29.5	1404	3	US-08-801-308-1
5	49	29.2	300	3	US-08-765-856-2
6	49	29.2	300	3	US-08-935-009A-2
7	49	29.2	302	3	US-08-765-856-4
8	49	29.2	302	3	US-08-935-009A-4
9	48.5	28.9	239	4	US-09-711-164-363
10	48.5	28.9	241	4	US-09-489-039A-8330
11	48.5	28.9	376	4	US-09-252-991A-24606
12	48	28.6	366	4	US-09-252-991A-31958
13	47	28.0	75	3	US-08-936-185A-506
14	47	28.0	312	3	US-08-910-501-2
15	47	28.0	312	3	US-08-910-501-4
16	47	28.0	312	3	US-09-398-550-2
17	47	28.0	312	3	US-09-398-550-4
18	47	28.0	1009	4	US-09-762-724-10
19	46.5	27.7	282	3	US-08-972-902-7
20	46.5	27.7	282	4	US-09-520-207-7
21	46.5	27.7	282	4	US-10-136-253-7
22	46.5	27.7	282	3	US-09-134-001C-5100
23	46	27.4	334	4	US-09-489-039A-10307
24	46	27.4	484	4	US-09-252-991A-29889
25	46	27.4	488	4	US-09-248-796A-15599
26	46	27.4	614	4	US-09-328-352-4504
27	45.5	27.1	137	4	US-10-101-464A-829

28	45.5	27.1	409	3	US-09-587-066-2	Sequence 2, Appli
29	45	26.8	252	4	US-09-270-767-43123	Sequence 43123, A
30	45	26.8	314	4	US-09-593-110-2807	Sequence 2807, Ap
31	45	26.8	356	4	US-09-248-796A-18231	Sequence 18231, A
32	45	26.8	421	4	US-09-248-796A-18642	Sequence 18642, A
33	45	26.8	475	3	US-09-491-362-2	Sequence 2, Appli
34	45	26.8	475	3	US-09-874-562-2	Sequence 2, Appli
35	45	26.8	490	4	US-09-252-991A-20767	Sequence 20767, A
36	45	26.8	523	4	US-09-107-532A-4149	Sequence 4149, Ap
37	45	26.8	526	4	US-08-311-731A-231	Sequence 231, App
38	45	26.8	538	4	US-08-311-731A-71	Sequence 71, Appl
39	45	26.8	1588	1	US-08-698-551-16	Sequence 16, Appl
40	45	26.8	1588	2	US-08-602-228-16	Sequence 16, Appl
41	45	26.8	1588	2	US-08-839-032A-16	Sequence 16, Appl
42	45	26.8	1588	3	US-09-185-258C-16	Sequence 16, Appl
43	45	26.8	1588	4	US-09-976-594-965	Sequence 965, Appl
44	44.5	26.5	185	3	US-09-122-443-16	Sequence 16, Appl
45	44.5	26.5	185	4	US-09-558-089-16	Sequence 16, Appl

ALIGNMENTS

RESULT 1
US-09-270-767-42036
; Sequence 42036, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 42036
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-42036

Query Match 31.0%; Score 52; DB 4; Length 510;
Best Local Similarity 36.7%; Pred. No. 8.4;
Matches 11; Conservative 6; Mismatches 5; Indels 8; Gaps 1;
Qy 4 VKDVLILSALRMLWAADFLDLPEQIG 33
Db 18 VKSIVVLAALTALLWA-----PLEVLG 39

RESULT 2
US-09-248-796A-14421
; Sequence 14421, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14421
; LENGTH: 491
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14421

Query Match 29.8%; Score 50; DB 4; Length 491;

Best Local Similarity 35.5%; Pred. No. 17;
Matches 11; Conservative 10; Mismatches 8; Indels 2; Gaps 1;
QY 4 VKDVLILSALRM--LWAADDFLEDLPEQI 32
DB 291 VKSVLQLLLIQLSLFESDDFYENVYDYL 321

RESULT 3
US-09-134-000C-6601
; Sequence 6601, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/955,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6601
; LENGTH: 306
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6601

Query Match 29.5%; Score 49.5; DB 4; Length 306;
Best Local Similarity 40.7%; Pred. No. 11;
Matches 11; Conservative 8; Mismatches 5; Indels 3; Gaps 1;
QY 6 DVLILSALRM--LWAADDFLEDLPEQI 32
DB 141 DTLVLAQAQISCAA---IEELPMQV 164

RESULT 4
US-08-801-308-1
; Sequence 1, Application US/08801308
; Patent No. 6368790
; GENERAL INFORMATION:
; APPLICANT: Scott, Robert E.
; TITLE OF INVENTION: CDNA ENCODING P2P PROTEINS AND USE OF
; TITLE OF INVENTION: P2P CDNA-DERIVED ANTIBODIES AND ANTISENSE REAGENTS IN
; DETERMINING THE PROLIFERATIVE POTENTIAL OF NORMAL,
; TITLE OF INVENTION: ABNORMAL AND CANCER CELLS IN ANIMALS AND HUMANS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Weiser & Associates, P.C.
; STREET: 230 S. Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/801,308
; FILING DATE: 18-FEB-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 372.6435P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-801-308-1

Query Match 29.5%; Score 49.5; DB 3; Length 1404;
Best Local Similarity 31.6%; Pred. No. 71;
Matches 12; Conservative 8; Mismatches 11; Indels 7; Gaps 2;
QY 3 PVKDVILSALRM--LWAADDFLEDL---PFEQIG 33
DB 1094 PAEDVIMIQVPSQKWKDFESEEDVKTQPIQSVG 1131

RESULT 5
US-08-765-856-2
; Sequence 2, Application US/08765856
; Patent No. 6074817
; GENERAL INFORMATION:
; APPLICANT: Landini, Maria P.
; APPLICANT: Ripalti, Alessandro
; APPLICANT: Maine, Gregory T.
; APPLICANT: Flanders, Richard T.
; TITLE OF INVENTION: RECOMBINANT MONO AND POLY ANTIGENS TO DETECT CYTOMEGALOVIRUS-SI
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,856
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/IT95/00073
; FILING DATE: 15-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Weinstein, David L.
; REGISTRATION NUMBER: 28,128
; REFERENCE/DOCKET NUMBER: 5750.US.O1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 937-6182
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 300 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-765-856-2

Query Match 29.2%; Score 49; DB 3; Length 300;
Best Local Similarity 38.5%; Pred. No. 13;
Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;
QY 6 DVLILSALRM--LWAADDFLEDLPEQ 31
DB 48 DTLVYASRNGLEFAVENFLTEPFQR 73

RESULT 6
US-08-935-009A-2

```
; Sequence 2, Application US/08935009A
; Patent No. 6177241
; GENERAL INFORMATION:
; APPLICANT: Maine, Gregory T.
; TITLE OF INVENTION: USE OF PEPTIDES TO IMPROVE
; TITLE OF INVENTION: SPECIFICITY OF AN ENZYME IMMUNOASSAY FOR
; TITLE OF INVENTION: THE DETECTION OF HERPESVIRUS SPECIFIC IGM ANTIBODY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/935.009A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weinstein, David L.
; REGISTRATION NUMBER: 28,128
; REFERENCE/DOCKET NUMBER: 6186.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-937-6182
; TELEFAX: 847-938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 300 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-935-009A-2
;
; Query Match 29.2%; Score 49; DB 3; Length 300;
; Best Local Similarity 38.5%; Pred. No. 13;
; Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;
;
; Qy 6 DVLLSALRRLMWADDFLEDPFEQ 31
; Db 48 DTLVVASRNGLFAVENFLTEEPQR 73
;
; RESULT 7
; US-08-765-856-4
; Sequence 4, Application US/08765856
; Patent No. 6074817
; GENERAL INFORMATION:
; APPLICANT: Landini, Maria P.
; APPLICANT: Ripalti, Alessandro
; APPLICANT: Maine, Gregory T.
; APPLICANT: Planders, Richard T.
; TITLE OF INVENTION: RECOMBINANT MONO AND POLY ANTIGENS TO DETECT CYTOMEGALOVIRUS-S
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/765,856
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/IT95/00073
; FILING DATE: 15-MAY-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Weinstein, David L.
; REGISTRATION NUMBER: 28,128
; REFERENCE/DOCKET NUMBER: 5750.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (847) 937-6182
; TELEFAX: (847) 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-765-856-4
;
; Query Match 29.2%; Score 49; DB 3; Length 302;
; Best Local Similarity 38.5%; Pred. No. 13;
; Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;
;
; Qy 6 DVLLSALRRLMWADDFLEDPFEQ 31
; Db 118 DTLVVASRNGLFAVENFLTEEPQR 143
;
; RESULT 8
; US-08-935-009A-4
; Sequence 4, Application US/08935009A
; Patent No. 6177241
; GENERAL INFORMATION:
; APPLICANT: Maine, Gregory T.
; TITLE OF INVENTION: USE OF PEPTIDES TO IMPROVE
; TITLE OF INVENTION: SPECIFICITY OF AN ENZYME IMMUNOASSAY FOR
; TITLE OF INVENTION: THE DETECTION OF HERPESVIRUS SPECIFIC IGM ANTIBODY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/935,009A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Weinstein, David L.
; REGISTRATION NUMBER: 28,128
; REFERENCE/DOCKET NUMBER: 6186.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847-937-6182
; TELEFAX: 847-938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-935-009A-4

Query Match      29.2%; Score 49; DB 3; Length 302;
Best Local Similarity 38.5%; Pred. No. 13;
Matches 10; Conservative 7; Mismatches 9; Indels 0; Gaps 0;

QY      6 DVILSALRRMLWAADDFLEDLPFEQ 31
DB     118 DFLLYVASRNGLFVAVENTLPEPFR 143

RESULT 9
US-09-711-164-363
; Sequence 363, Application US/09711164
; Patent No. 6589738
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-9
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 363
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-711-164-363

Query Match      28.9%; Score 48.5; DB 4; Length 239;
Best Local Similarity 38.2%; Pred. No. 12;
Matches 13; Conservative 6; Mismatches 8; Indels 7; Gaps 2;

QY      7 VILSAL---RRMLWAAD---DFLEDLPFEQIG 33
DB     15 LLLLTVLGLDRWWSKMTAPYYDELQDLFQRQVG 48

RESULT 10
US-09-489-039A-8330
; Sequence 8330, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709-2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 8330
; LENGTH: 241
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-8330

Query Match      28.9%; Score 48.5; DB 4; Length 241;
Best Local Similarity 38.2%; Pred. No. 12;
Matches 13; Conservative 6; Mismatches 8; Indels 7; Gaps 2;

QY      7 VILSAL---RRMLWAAD---DFLEDLPFEQIG 33
DB     18 LLLLTVLGLDRWWSKMTAPYYDELQDLFQRQVG 51
```

```
RESULT 11
US-09-252-991A-24606
; Sequence 24606, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24606
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24606
```

```
Query Match      28.9%; Score 48.5; DB 4; Length 376;
Best Local Similarity 40.0%; Pred. No. 21;
Matches 12; Conservative 3; Mismatches 14; Indels 1; Gaps 1;
```

```
QY      3 PVKDVILSAL-RRMLWAADDFLEDLPFEQ 31
DB     70 PVKGLYFVGVGGRGKTYLVDTFESLPFEQ 99
```

```
RESULT 12
US-09-252-991A-31958
; Sequence 31958, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31958
; LENGTH: 366
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31958
```

```
Query Match      28.6%; Score 48; DB 4; Length 366;
Best Local Similarity 40.0%; Pred. No. 24;
Matches 12; Conservative 3; Mismatches 5; Indels 10; Gaps 1;
```

```
QY      3 PVKDVILSALRRMLWAADDFLEDLPFEQ 32
DB     198 PVNDLALL-----DPRLLDLPFEQ 217
```

```
RESULT 13
US-08-936-165A-506
; Sequence 506, Application US/08936165A
; Patent No. 6348582
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999

```

```

; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 312 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-910-501-4

Query Match      28.0%; Score 47; DB 3; Length 312;
Best Local Similarity 81.8%; Pred. No. 29;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      22 DFLEDLPPFQI 32
      |||: |||||
Db      162 DELKTLPPFQI 172

```

Search completed: November 10, 2004, 13:44:00
Job time : 12.399 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 16:36:12 ; Search time 35.4508 Seconds
(without alignments)
328.807 Million cell updates/sec

Title: US-10-092-750-25

Perfect score: 168
Sequence: 1 LDPVKDVLILSALRMLWAADDFLEDLPFEQIG 33

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1566620 seqs, 353225886 residues

Total number of hits satisfying chosen parameters: 1566620

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	168	100.0	33	14	US-10-092-750-25
2	168	100.0	86	14	Sequence 25, Appl
3	168	100.0	444	16	Sequence 5837, Ap
4	132	78.6	413	14	Sequence 220, App
5	55	32.7	473	17	Sequence 141, App
6	54	32.1	67	15	Sequence 6676, Ap
7	54	32.1	346	15	Sequence 260441, A
8	51	30.4	481	15	Sequence 259025, A
9	50	29.8	1839	14	Sequence 48632, A
10	49.5	29.5	297	9	Sequence 7605, Ap
11	49.5	29.5	297	15	Sequence 10933, A
12	49.5	29.5	1030	16	Sequence 56875, A
13	49.5	29.5	1242	15	Sequence 141820, A
					Sequence 70111, A

Sequence 1, Appli
Sequence 69906, A
Sequence 3594, Ap
Sequence 3608, Ap
Sequence 113343,
Sequence 3, Appli
Sequence 29, Appl
Sequence 1608, Ap
Sequence 14, Appl
Sequence 21, Appl
Sequence 174, App
Sequence 203003,
Sequence 138446,
Sequence 26, Appl
Sequence 363, App
Sequence 145784,
Sequence 18637, A
Sequence 146, App
Sequence 9183, Ap
Sequence 151727,
Sequence 164301,
Sequence 164299,
Sequence 228991,
Sequence 62689, A
Sequence 204511,
Sequence 506, App
Sequence 18281, A
Sequence 10, Appl
Sequence 7, Appl
Sequence 5553, Ap
Sequence 355182,
Sequence 174465,

ALIGNMENTS

RESULT 1

US-10-092-750-25
; Sequence 25, Application US/10092750
; Publication No. US20030032157A1
; GENERAL INFORMATION:
; APPLICANT: Hammond, Philip W.
; APPLICANT: Alpin, Julia
; APPLICANT: Wright, Martin C.
; TITLE OF INVENTION: Polypeptides Interactive with BCL-X1
; FILE REFERENCE: 50036/050002
; CURRENT APPLICATION NUMBER: US/10/092,750
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: US 60/274,526
; NUMBER OF SEQ ID NOS: 253
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 33
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-750-25

Query Match 100.0%; Score 168; DB 14; Length 33;
Best Local Similarity 100.0%; Pred. No. 1.8e-17;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LDPVKDVLILSALRMLWAADDFLEDLPFEQIG 33

Db 1 LDPVKDVLILSALRMLWAADDFLEDLPFEQIG 33

RESULT 2

US-10-106-698-5837
; Sequence 5837, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:

```
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; PRIOR FILING DATE: 1999-11-03
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5837
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5837

Query Match 100.0%; Score 168; DB 14; Length 86;
Best Local Similarity 100.0%; Pred. No. 5,1e-17;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LDPVKDVLILSALRMLWAADDFLEDLPFEQIG 33
Db 40 LDPVKDVLILSALRMLWAADDFLEDLPFEQIG 72

RESULT 3
US-10-755-889-220
; Sequence 220, Application US/10755889
; Publication No. US20040171823A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB
; FILE REFERENCE: D0284 NP
; CURRENT APPLICATION NUMBER: US/10/755,889
; CURRENT FILING DATE: 2004-01-13
; PRIOR APPLICATION NUMBER: U.S. 60/440,068
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: U.S. 60/469,757
; PRIOR FILING DATE: 2003-05-12
; NUMBER OF SEQ ID NOS: 823
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 220
; LENGTH: 444
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-755-889-220

Query Match 100.0%; Score 168; DB 16; Length 444;
Best Local Similarity 100.0%; Pred. No. 3.2e-16;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LDPVKDVLILSALRMLWAADDFLEDLPFEQIG 33
Db 398 LDPVKDVLILSALRMLWAADDFLEDLPFEQIG 430

RESULT 4
US-10-097-340-141
; Sequence 141, Application US/10097340
; Publication No. US20030087250A1
; GENERAL INFORMATION:
; APPLICANT: John MONAHAN
; APPLICANT: Manjula GANNAVAPURU
; APPLICANT: Sebastian HOERSCH
; APPLICANT: Shubhangi KAMATKAR
; APPLICANT: Steve G. KOVATS
; APPLICANT: Rachel E. MEYERS
; APPLICANT: Michael MORRISBY
; APPLICANT: Peter OLANDT
```

```
; APPLICANT: Ami SEN
; APPLICANT: Peter VEIBY
; APPLICANT: Gordon B. MILLS
; APPLICANT: Robert C. EAST, Jr.
; APPLICANT: Karen LU
; APPLICANT: Rosemarie SCHMANDT
; APPLICANT: Xumei ZHAO
; APPLICANT: Karen GLATT
; TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,
; TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer
; FILE REFERENCE: MRI-030
; CURRENT APPLICATION NUMBER: US/10/097,340
; CURRENT FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 60/276,025
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/325,149
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/276,026
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/324,967
; PRIOR FILING DATE: 2001/09/26
; PRIOR APPLICATION NUMBER: 60/311,732
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/325,102
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/323,580
; PRIOR FILING DATE: 2001-09-19
; NUMBER OF SEQ ID NOS: 363
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 141
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-097-340-141

Query Match 78.6%; Score 132; DB 14; Length 413;
Best Local Similarity 78.8%; Pred. No. 6.3e-11;
Matches 26; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LDPVKDVLILSALRMLWAADDFLEDLPFEQIG 33
Db 371 LDPVKDVLILSALRMLWAADDFLEDLPFEQIG 403

RESULT 5
US-10-739-930-6676
; Sequence 6676, Application US/10739930
; Publication No. US20040216190A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES AND OTHER MOLECULES ASSOCIATED WITH
; TITLE OF INVENTION: PLANTS AND USES THEREOF FOR PLANT IMPROVEMENT
; FILE REFERENCE: 38-21(5337)B
; CURRENT APPLICATION NUMBER: US/10/739,930
; CURRENT FILING DATE: 2003-12-18
; NUMBER OF SEQ ID NOS: 11088
; SEQ ID NO 6676
; LENGTH: 473
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: Clone ID: ARATH-23APR03-C6869_1.P
US-10-739-930-6676

Query Match 32.7%; Score 55; DB 17; Length 473;
Best Local Similarity 41.9%; Pred. No. 18;
Matches 13; Conservative 6; Mismatches 10; Indels 2; Gaps 1;

QY 4 VKDVLILSALRMLWAADDFLEDLPFEQIG 32
Db 177 VKSVYGLIQKGVFWAFQDMQIETGDLPEEI 207
```


RESULT 6

US-10-424-599-260441
 ; Sequence 260441, Application US/10424599
 ; Publication No. US20040031072A1

GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 260441

; LENGTH: 67

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_77200C.1.pep

US-10-424-599-260441

Query Match 32.1%; Score 54; DB 15; Length 67;

Best Local Similarity 45.7%; Pred. No. 2.9;

Matches 16; Conservative 3; Mismatches 14; Indels 2; Gaps 1;

Qy 1 LDPVKDVLILSALRRMLWAADDFLE--DLPEQIG 33

Db 23 LDPEKPSVLDKLRMLAEADAALAEARKKPLEEKG 57

RESULT 7

US-10-424-599-259025

; Sequence 259025, Application US/10424599

; Publication No. US20040031072A1

GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 259025

; LENGTH: 346

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_75923C.1.pep

US-10-424-599-259025

Query Match 32.1%; Score 54; DB 15; Length 346;

Best Local Similarity 45.7%; Pred. No. 18;

Matches 16; Conservative 3; Mismatches 14; Indels 2; Gaps 1;

Qy 1 LDPVKDVLILSALRRMLWAADDFLE--DLPEQIG 33

Db 280 LDPEKPSVLDKLRMLAEADAALAEARKKPLEEKG 314

RESULT 8

US-10-282-122A-48632

; Sequence 48632, Application US/10282122A

; Publication No. US20040029129A1

GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu
 ; APPLICANT: Zamudio, Carlos
 ; APPLICANT: Malone, Cheryl
 ; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Wall, Daniel
 ; APPLICANT: Trawick, John
 ; APPLICANT: Carr, Grant
 ; APPLICANT: Yamamoto, Robert
 ; APPLICANT: Forsyth, R.
 ; APPLICANT: Xu, H.
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
 ; FILE REFERENCE: ELITRA.034A
 ; CURRENT APPLICATION NUMBER: US/10/282,122A
 ; CURRENT FILING DATE: 2003-02-20
 ; PRIOR APPLICATION NUMBER: 60/191,078
 ; PRIOR FILING DATE: 2000-03-21
 ; PRIOR APPLICATION NUMBER: 60/206,848
 ; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/230,335
 ; PRIOR FILING DATE: 2000-09-06
 ; PRIOR APPLICATION NUMBER: 60/230,347
 ; PRIOR FILING DATE: 2000-09-09
 ; PRIOR APPLICATION NUMBER: 60/242,578
 ; PRIOR FILING DATE: 2000-10-23
 ; PRIOR APPLICATION NUMBER: 60/253,625
 ; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931
 ; PRIOR FILING DATE: 2000-12-22
 ; PRIOR APPLICATION NUMBER: 60/267,636
 ; PRIOR FILING DATE: 2001-02-09
 ; PRIOR APPLICATION NUMBER: 60/269,308
 ; PRIOR FILING DATE: 2001-02-16
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; SOFTWARE: PatentIn version 3.1
 ; NUMBER OF SEQ ID NOS: 78614
 ; SEQ ID NO 48632
 ; LENGTH: 481
 ; TYPE: PRT
 ; ORGANISM: Bacteroides fragilis
 ; US-10-282-122A-48632

Query Match 30.4%; Score 51; DB 15; Length 481;

Best Local Similarity 50.0%; Pred. No. 73;

Matches 11; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 6 DVLILSALRRMLWAADDFLEDL 27

Db 353 DVVTSAVRLLFEAGDDIDDL 374

RESULT 9

US-10-032-585-7605

; Sequence 7605, Application US/10032585

; Publication No. US20030180953A1

GENERAL INFORMATION:

; APPLICANT: Terry, Roemer D.
 ; APPLICANT: Bo, Jiang
 ; APPLICANT: Charles, Boone
 ; APPLICANT: Howard, Bussey

; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery

; FILE REFERENCE: 10182-005-999

; CURRENT APPLICATION NUMBER: US/10/032,585

; CURRENT FILING DATE: 2001-12-20

; NUMBER OF SEQ ID NOS: 8000

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 7605

; LENGTH: 1839

; TYPE: PRT

; ORGANISM: Candida albicans

US-10-032-585-7605

Query Match 29.8%; Score 50; DB 14; Length 1839;

Best Local Similarity 35.5%; Pred. No. 4.5e+02;

QY 1 LDPVKDVLILSALRRM-----LWAADDF-----LEDLPFFQIG 33
Db 726 VDGIKVITASALRRRTDRAKAEDPDSBEDLLRENECEDEIFQIG 775

RESULT 13

US-10-282-122A-70111
; Sequence 70111, Application US/10282122A
; Publication No. US20040029129A1

GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIORITY FILING DATE: 2003-02-20

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-09-06

; PRIOR FILING DATE: 2000-09-06

; PRIOR FILING DATE: 2000-09-09

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2001-02-09

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 70111

; LENGTH: 1242

; TYPE: PRT

; ORGANISM: Pseudomonas syringae

US-10-282-122A-70111

Query Match 29.5%; Score 49.5; DB 15; Length 1242;

Best Local Similarity 42.3%; Pred. No. 3.5e+02;

Matches 11; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

QY 7 VLILSALRRMLWAADDFLEDLPFEQI 32

Db 301 VTLARIKARVAAQDH-QDLPPFEQV 325

RESULT 14

US-09-811-045A-1

; Sequence 1, Application US/09811045A

; Patent No. US20020035080A1

GENERAL INFORMATION:

; APPLICANT: Scott, Robert E.

; TITLE OF INVENTION: cDNA encoding P2P proteins and use of P2P cDNA-

; TITLE OF INVENTION: derived antibodies and antisense reagents

; TITLE OF INVENTION: in determining the proliferative potential of

; TITLE OF INVENTION: normal, abnormal and cancer cells in animals
; FILE REFERENCE: D6386D

; CURRENT APPLICATION NUMBER: US/09/811,045A

; PRIORITY FILING DATE: 2001-03-16

; PRIOR FILING DATE: 1997-02-18

; NUMBER OF SEQ ID NOS: 4

; SEQ ID NO 1

; LENGTH: 1404

; TYPE: PRT

; ORGANISM: Unknown

; FEATURE:

; NAME/KEY: PEPTIDE

; OTHER INFORMATION: P2P polypeptide

US-09-811-045A-1

Query Match 29.5%; Score 49.5; DB 9; Length 1404;

Best Local Similarity 31.6%; Pred. No. 4e+02;

Matches 12; Conservative 8; Mismatches 11; Indels 7; Gaps 2;

QY 3 PVKDVILILSALRRMLWAADDF---LEDL---PFFQIG 33

Db 1094 PAEDVILMIQVPSKWDKDDFESEEDVKTTQPIQSVG 1131

RESULT 15

US-10-282-122A-69906

; Sequence 69906, Application US/10282122A

; Publication No. US20040029129A1

GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu

; APPLICANT: Zamudio, Carlos

; APPLICANT: Malone, Cheryl

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari

; APPLICANT: Zyskind, Judith

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John

; APPLICANT: Carr, Grant

; APPLICANT: Yamamoto, Robert

; APPLICANT: Forsyth, R.

; APPLICANT: Xu, H.

; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

; FILE REFERENCE: ELITRA.034A

; CURRENT APPLICATION NUMBER: US/10/282,122A

; PRIORITY FILING DATE: 2003-02-20

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-23

; PRIOR FILING DATE: 2000-05-26

; PRIOR FILING DATE: 2000-09-06

; PRIOR FILING DATE: 2000-09-09

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-11-27

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2000-12-22

; PRIOR FILING DATE: 2001-02-09

; PRIOR FILING DATE: 2001-02-16

; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 69906

; LENGTH: 323

; TYPE: PRT

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:29:32 ; Search time 7.35233 Seconds
(without alignments)
431.857 Million cell updates/sec

Title: US-10-092-750-25
Perfect score: 168
Sequence: 1 LDPVKVLLSALRMLWAADDFLEDPFEQIG 33

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: Pir1.*
2: Pir2.*
3: Pir3.*
4: Pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	168	100.0	444	2 A43676	P44 hepatitis-asso
2	168	100.0	444	2 S48218	microtubular aggre
3	55	32.7	473	2 D96591	hypothetical prote
4	54.5	32.4	378	2 E84806	probable elongatio
5	54	32.1	514	2 T39217	centromere k-type
6	54	32.1	2420	2 A84652	hypothetical prote
7	53	31.5	178	2 A45540	replicase protein
8	53	31.5	1839	1 ERWPEM	genome polyprotein
9	52	31.0	209	2 G72341	uracil phosphoribo
10	51	30.4	411	2 B84504	probable retroelem
11	50	29.8	138	2 JC6308	glia maturation fa
12	50	29.8	354	2 H85860	O6-methylguanine-D
13	50	29.8	354	2 F91016	threonine synthase
14	50	29.8	514	1 SYBSR	gene NDI intron 3
15	50	29.8	580	2 S06057	hypothetical prote
16	49.5	29.5	416	2 T26371	H+-transporting tw
17	49.5	29.5	502	2 S01401	hypothetical prote
18	49	29.2	109	2 B86050	hypothetical prote
19	49	29.2	141	2 H91203	hypothetical prote
20	49	29.2	323	2 T42683	hypothetical prote
21	49	29.2	354	1 XVECO2	methylated-DNA-lpr
22	49	29.2	433	1 QQBEV2	UL44 protein - hum
23	49	29.2	470	2 A65168	hypothetical 49.9
24	49	29.2	636	2 AH1889	hypothetical prote
25	49	29.2	1076	2 B96682	protein Flr22.14 (
26	48.5	28.9	196	2 A11301	hypothetical prote
27	48.5	28.9	198	2 I39732	hypothetical prote
28	48.5	28.9	239	1 G64982	sana protein Esc
29	48.5	28.9	239	2 AF0780	vancomycin resista

ALIGNMENTS

RESULT 1

A43676
P44 hepatitis-associated antigen - chimpanzee
C:Species: Pan troglodytes (chimpanzee)
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 07-May-1999
C:Accession: A43676; A43677
R:Takahashi, K.; Kitamura, N.; Shibui, T.; Kamizono, M.; Matsui, R.; Yoshiyama, Y.; Maeda
J. Gen. Virol. 71, 2005-2011, 1990
A:Title: Cloning, sequencing and expression in Escherichia coli of cDNA for a non-A, non-
A:Reference number: A43676; MUID:91011346; PMID:2170570
A:Accession: A43676
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-444 <TAK>
A:Cross-references: GB:D90034
R:Honda, Y.; Kondo, J.; Maeda, T.; Yoshiyama, Y.; Yamada, E.; Shimizu, Y.K.; Shikata, T.;
J. Gen. Virol. 71, 1999-2004, 1990
A:Title: Isolation and purification of a non-A, non-B hepatitis-associated microtubular
A:Reference number: A43677; MUID:91011345; PMID:2170569
A:Accession: A43677
A:Status: preliminary
A:Molecule type: protein
A:Residues: 14-22;108-128;200-214;231-240 <HON>
Query Match 100.0%; Score 168; DB 2; Length 444;
Best Local Similarity 100.0%; Pred. No. 1.6e-16;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LDPVKVLLSALRMLWAADDFLEDPFEQIG 33
DB 398 LDPVKVLLSALRMLWAADDFLEDPFEQIG 430

RESULT 2

S48218
microtubular aggregate protein - human
C:Species: Homo sapiens (man)
C:Date: 15-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 19-Oct-1995
C:Accession: S48218
R:Kitamura, A.; Takahashi, K.; Okajima, A.; Kitamura, N.
Eur. J. Biochem. 224, 877-883, 1994
A:Title: Induction of the human gene for p44, a hepatitis-C-associated microtubular aggr
A:Reference number: S48218; MUID:95010078; PMID:7925411
A:Accession: S48218
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-444 <KIT>
Query Match 100.0%; Score 168; DB 2; Length 444;
Best Local Similarity 100.0%; Pred. No. 1.6e-16;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LDPVKDVLILSALRMLWAADDFLEDLPPFEQIG 33
 |||||
 Db 398 LDPVKDVLILSALRMLWAADDFLEDLPPFEQIG 430
 |||||

RESULT 3
 D96591
 hypothetical protein T24C10.10 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
 C:Accession: D96591
 R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 C:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziali,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: D96591
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-473 <STO>
 A:Cross-references: UNIPROT:Q9P233; GB:AE005173; NID:G9857523; PIDN:AAG00878.1; GSPDB:GN
 C:Genetics:
 A:Gene: T24C10.10
 A:Map position: 1

Query Match 32.7%; Score 55; DB 2; Length 473;
 Best Local Similarity 41.9%; Pred. No. 5;
 Matches 13; Conservative 6; Mismatches 10; Indels 2; Gaps 1;

QY 4 VKDVLILSALRMLWAADDFLE--DLPPFEQI 32
 |||||
 Db 177 VKEVGLTQEGVFWAFQMTGTGLPYEEI 207
 |||||

RESULT 4
 E84806
 probable elongation factor [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
 C:Accession: E84806
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: E84806
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2420 <STO>
 A:Cross-references: UNIPROT:Q9SL90; GB:AE002093; NID:G4874311; PIDN:AAD31373.1; GSPDB:GN
 C:Genetics:
 A:Gene: At2G25730
 A:Map position: 2

Query Match 32.1%; Score 54; DB 2; Length 2420;
 Best Local Similarity 50.0%; Pred. No. 42;
 Matches 12; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 2 DPVKDVLILSALRMLWAADDFLE 25
 |||||
 Db 861 DQVHEVALISVLRRLIKSLQEFLE 884
 |||||

RESULT 7
 A45540
 replicase protein - Physalis mottle virus (fragment)
 C:Species: Physalis mottle virus
 C:Date: 22-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
 C:Accession: A45540
 R:Jacob, A.N.K.; Murthy, M.R.; Savithri, H.S.
 Arch. Virol. 123, 367-377, 1992
 A:Title: Nucleotide sequence of the 3' terminal region of belladonna mottle virus-Iowa (1
 A:Reference number: A45540; MUID:92222369; PMID:1562236
 A:Accession: A45540
 A:Status: preliminary
 A:Molecule type: Genomic RNA
 A:Residues: 1-178 <JAC>
 A:Cross-references: UNIPROT:P36352; GB:S97776; NID:G248382; PIDN:AAB21996.1; PID:G248383
 A:Note: sequence extracted from NCBI backbone (NCBI:97776, NCBIP:97778)
 A:Note: in Genbank entry S97776, release 109.0, the source is designated as Belladonna mc

Query Match 32.4%; Score 54.5; DB 2; Length 378;
 Best Local Similarity 31.4%; Pred. No. 4.6;
 Matches 16; Conservative 8; Mismatches 8; Indels 19; Gaps 2;

QY 2 DPVKD---VLILSALRMLWAADDFLE-----LEDLPPFEQIG 33
 |||||
 Db 208 DPVRDKIRELLVEALCRVAGEADYRESVNASDPLRVAVSVESLMFEKIG 258
 |||||

RESULT 5
 T39217
 centromere k-type repeat binding protein Cbhp - fission yeast (Schizosaccharomyces pombe

Qy 5 KDVILSALRRMLWAADDLELPFEQIG 33
|||::||| ||| ||| ||| :|
Db 70 KDIVVPILRAGLVMA DGILELLPNASVG 98

```
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ada
C:Superfamily: adaptive response regulatory protein; methylated-DNA-protein-cysteine S-P
C:Keywords: methylated amino acid
F:321/Binding site: methyl (Cys) (covalent) #status predicted

Query Match      29.8%; Score 50; DB 2; Length 354;
Best Local Similarity 42.3%; Pred. No. 19;
Matches 11; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 6 DVILSALRMLWAADDFLEDLPEEQ 31
Db 219 DATLISELQMPFAADNAPADLTFOQ 244

RESULT 13
F91016
O6-methylguanine-DNA methyltransferase [imported] - Escherichia coli (strain O157:H7, su
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: F91016
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F91016
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-354 <HAY>
A:Cross-references: UNIPROT:Q8XB42; GB:BA000007; PIDN:BA8362571; GSPDB:G
A:Experimental source: strain O157:H7, substrain RMD 050952
C:Genetics:
A:Gene: ECs3102
C:Superfamily: adaptive response regulatory protein; methylated-DNA-protein-cysteine S-P
C:Keywords: methylated amino acid
F:321/Binding site: methyl (Cys) (covalent) #status predicted

Query Match      29.8%; Score 50; DB 2; Length 354;
Best Local Similarity 42.3%; Pred. No. 19;
Matches 11; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 6 DVILSALRMLWAADDFLEDLPEEQ 31
Db 219 DATLISELQMPFAADNAPADLTFOQ 244

RESULT 14
SYBSR
threonine synthase (EC 4.2.3.1) - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YCR053w
C:Species: Saccharomyces cerevisiae
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 09-Jul-2004
C:Accession: S22836; S20154; S19467
R:Mannhaupt, G.; van der Linden, G.; Vetter, I.; Maurer, K.; Pilz, U.; Planta, R.; Feldm
Yeast 6, 353-361, 1990
A:Title: Analysis of the THR4 region on chromosome III of the yeast Saccharomyces cerevi
A:Reference number: S22836; MUID:90371958; PMID:2204248
A:Accession: S22836
A:Molecule type: DNA
A:Residues: 1-514 <MAN>
A:Cross-references: UNIPROT:P16120; EMBL:X59720; NID:g1907116; PIDN:CAA42284.1; PID:g190
R:Aas, S.F.; Rognes, S.E.
Nucleic Acids Res. 18, 665, 1990
A:Title: Nucleotide sequence of the yeast THR4 gene encoding threonine synthase.
A:Reference number: S20154; MUID:90175003; PMID:2408022
A:Accession: S20154
A:Molecule type: DNA
A:Residues: 1-514 <AAS>
A:Cross-references: EMBL:X17256; NID:g4615; PIDN:CAA35157.1; PID:g4616
R:van der Linden, C.G.; Maurer, C.T.C.; Planta, R.J.; van Vliet-Reedijk, J.C.; Vreken, F
submitted to the Protein Sequence Database, March 1992
```

```
A:Reference number: S19415
A:Accession: S19467
A:Molecule type: DNA
A:Residues: 1-514 <LIN>
A:Cross-references: EMBL:X59720; NID:g1907116; PIDN:CAA42284.1; PID:g1907196; GSPDB:GN000
C:Genetics:
A:Gene: SGD:THR4; MIPS:YCR053w
A:Cross-references: SGD:S0000649; MIPS:YCR053w
A:Map position: 3R
C:Superfamily: threonine synthase
C:Keywords: carbon-oxygen lyase; phosphoprotein; pyridoxal phosphate; threonine biosynth
F:124/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match      29.8%; Score 50; DB 1; Length 514;
Best Local Similarity 38.9%; Pred. No. 23;
Matches 14; Conservative 6; Mismatches 12; Indels 4; Gaps 2;

QY 1 LDPVKDVLILSRRLMW-AADDFL---EDLPFEQI 32
Db 327 LSPAMDILISSFERLLMYLALEYLANGDDLKAGEI 362

RESULT 15
S06057
gene NDI intron 3 protein 2 - Podospora anserina mitochondrion
C:Species: mitochondrion Podospora anserina
C:Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 09-Jul-2004
C:Accession: S06057
R:Cummings, D.J.; Domenico, J.M.; Michel, F.
Curr. Genet. 14, 253-264, 1988
A:Title: DNA sequence and organization of the mitochondrial NDI gene from Podospora anse
A:Reference number: S06056; MUID:89063443; PMID:3197134
A:Accession: S06057
A:Molecule type: DNA
A:Residues: 1-580 <CUM>
A:Cross-references: UNIPROT:Q02714; EMBL:X13164
C:Genetics:
A:Genome: mitochondrion
A:Genetic code: SGC3
C:Keywords: mitochondrion

Query Match      29.8%; Score 50; DB 2; Length 580;
Best Local Similarity 35.5%; Pred. No. 33;
Matches 11; Conservative 7; Mismatches 9; Indels 4; Gaps 1;

QY 4 VKDVLILSRRLMWAD---DFLEDLPPE 30
Db 350 LKDVLVSNIAATKWAYDALTLDLKSGPFE 380

Search completed: November 10, 2004, 13:40:35
Job time : 8.35233 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 10, 2004, 12:27:34 ; Search time 40.2953 Seconds
(without alignments)
471.205 Million cell updates/sec

Title: US-10-092-750-25

Perfect score: 168

Sequence: 1 LDPVKVILSALRMLWAADFLDLFPFQIG 33

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 02:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	168	100.0	444	1 IF44 HUMAN	Q8tcb0 homo sapien
2	168	100.0	444	1 IF44 HUMAN	P27473 pan troglod
3	168	100.0	444	2 Q14436	Q14496 homo sapien
4	132	78.6	413	2 Q99984	Q99984 homo sapien
5	132	78.6	413	2 Q96B64	Q96B64 homo sapien
6	110	65.5	396	2 Q8SP46	Q8SP46 mus musculus
7	106	63.1	422	2 Q8R0G4	Q8R0G4 mus musculus
8	106	63.1	422	2 Q8BV66	Q8BV66 mus musculus
9	106	63.1	422	2 Q8ER37	Q8ER37 mus musculus
10	100	59.5	449	2 Q8EDB7	Q8EDB7 mus musculus
11	60	35.7	110	2 Q8Y224	Q8Y224 pagrus major
12	60	35.7	110	2 AAP20189	AAP20189 pagrus major
13	58	34.5	651	2 Q8T749	Q8T749 branchiostoma
14	56	33.3	390	2 Q6NM08	Q6NM08 arabidopsis
15	56	33.3	390	2 AAS76264	AAS76264 arabidopsis
16	55.5	33.0	378	2 Q8VX26	Q8VX26 arabidopsis
17	55	32.7	392	2 Q85707	Q85707 clostridium
18	55	32.7	441	2 Q94Q07	Q94Q07 arabidopsis
19	55	32.7	473	2 Q9F233	Q9F233 arabidopsis
20	54.5	32.4	378	2 Q9ZVH8	Q9ZVH8 arabidopsis
21	54	32.1	294	2 Q93561	Q93561 perca fluviatilis
22	54	32.1	341	2 Q89513	Q89513 andean puma
23	54	32.1	514	1 CBH1 SCHPO	O14423 schizosaccharomyces
24	54	32.1	716	2 Q7QB55	Q7QB55 anopheles gambiae
25	54	32.1	2464	2 Q9SL90	Q9SL90 arabidopsis
26	53	31.5	178	1 POLR_PPMV	P36352 physalis mollis
27	53	31.5	281	2 Q89D37	Q89D37 bradyrhizobium
28	53	31.5	322	2 Q72JY5	Q72JY5 thermus thermophilus
29	53	31.5	322	2 AAS80981	AAS80981 thermus thermophilus
30	53	31.5	1839	1 POLR_EPMV	P20126 eggplant
31	52	31.0	209	1 UDP_THEMEA	Q9w210 thermotoga maritima

32	52	31.0	738	2	Q9W366	Q9W366 drosophila
33	51	30.4	327	2	Q6D5P8	Q6D5P8 erwinia carotovora
34	51	30.4	402	2	Q8KE54	Q8KE54 chlorobium
35	51	30.4	411	2	Q9SKG3	Q9SKG3 arabidopsis
36	51	30.4	411	2	Q8S8B3	Q8S8B3 arabidopsis
37	51	30.4	502	2	Q8A6B0	Q8A6B0 bacteroides
38	51	30.4	540	2	Q965Q6	Q965Q6 caenorhabditis
39	51	30.4	1835	2	Q9QCX3	Q9QCX3 chavotea
40	50.5	30.1	406	2	Q9RDQ9	Q9RDQ9 streptomyces
41	50.5	30.1	736	2	Q6L417	Q6L417 solanum demissum
42	50.5	30.1	736	2	AAT39966	AAT39966 solanum demissum
43	50	29.8	130	2	Q8CASA9	Q8CASA9 mus musculus
44	50	29.8	131	2	Q9AIK8	Q9AIK8 pseudomonas
45	50	29.8	131	2	Q9AIK9	Q9AIK9 pseudomonas

ALIGNMENTS

RESULT 1

IP44 HUMAN	STANDARD;	PRT;	444 AA.
AC	Q8TCE0;		
DT	05-JUL-2004 (Rel. 44, Created)		
DT	05-JUL-2004 (Rel. 44, Last sequence update)		
DT	05-JUL-2004 (Rel. 44, Last annotation update)		
DE	Interferon-induced protein 44 [p44].		
GN	Name=IF144; Synonyms=MTAP44;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A., AND INDUCTION.		
RX	MEDLINE=95010078; PubMed=7925411;		
RA	Kitamura A., Takahashi K., Okajima A., Kitamura N.;		
RT	"Induction of the human gene for p44, a hepatitis-C-associated		
RT	microtubular aggregate protein, by interferon-alpha/beta.";		
RL	Eur. J. Biochem. 224:877-883 (1994).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Lung;		
RX	MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;		
RA	Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D.,		
RA	Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,		
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,		
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,		
RA	Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,		
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,		
RA	Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,		
RA	Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,		
RA	Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,		
RA	Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs S.W.,		
RA	Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,		
RA	Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,		
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,		
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,		
RA	Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,		
RA	Schneerch A., Schein J.E., Jones S.J.M., Marra M.A.;		
RT	"Generation and initial analysis of more than 15,000 full-length human		
RT	and mouse cDNA sequences.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).		
CC	-!- FUNCTION: This protein aggregates to form microtubular structures		
CC	(by similarity).		
CC	-!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).		
CC	-!- INDUCTION: By alpha and beta interferons, but not by gamma		
CC	interferons.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; BC022870; AAH22870.1; -
 DR Genew; HGNC:16938; IFI44.
 KW Interferon induction.
 SQ SEQUENCE 444 AA; 50461 MW; BC26CA9D7CE41F2F CRC64;

Query Match 100.0%; Score 168; DB 1; Length 444;
 Best Local Similarity 100.0%; Pred. No. 1.2e-15;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LDPVKDVLILSALRRMLWAADDFLEDLPFEQIG 33
 DB 398 LDPVKDVLILSALRRMLWAADDFLEDLPFEQIG 430

RESULT 2
 IF44 PANTR
 ID IF44 PANTR STANDARD; PRT; 444 AA.
 AC P27473;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Interferon-induced protein 44 (Antigen p44) (Non-A non-B hepatitis-
 DE associated microtubular aggregates protein).
 GN Name=IFI44;
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 14-22; 108-128 AND 231-240.
 RC TISSUE=Liver;
 RX MEDLINE=91011346; PubMed=2170570;
 RA Takahashi K., Kitamura N., Shibui T., Kamizono M., Matsui R.,
 RA Yoshiyama Y., Maeda T., Kondo J., Honda Y., Yamada E., Shimizu Y.K.,
 RA Teranishi Y., Nakanishi S.;
 RT "Cloning, sequencing and expression in Escherichia coli of cDNA for a
 RT non-A, non-B hepatitis-associated microtubular aggregates protein.";
 RL J. Gen. Virol. 71:2005-2011(1990).
 RN [2]

RP PARTIAL SEQUENCE.
 RC TISSUE=Liver;
 RX MEDLINE=91011345; PubMed=2170569;
 RA Honda Y., Kondo J., Maeda T., Yoshiyama Y., Yamada E., Shimizu Y.K.,
 RA Shikata T., Ono Y.;
 RT "Isolation and purification of a non-A, non-B hepatitis-associated
 RT microtubular aggregates protein.";
 RL J. Gen. Virol. 71:1999-2004(1990).
 CC - FUNCTION: This protein aggregates to form microtubular structures.
 CC - SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC - TISSUE SPECIFICITY: Hepatocytes.
 CC - INDUCTION: During infection of non-A non-B hepatitis virus or
 CC hepatitis delta virus.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; D90034; BAA14082.1; -
 DR Fram; F07034; TLD; 1.
 KW Antigen; Direct protein sequencing.
 SQ SEQUENCE 444 AA; 50471 MW; 799CC855018CFA71 CRC64;

Query Match 100.0%; Score 168; DB 1; Length 444;
 Best Local Similarity 100.0%; Pred. No. 1.2e-15;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LDPVKDVLILSALRRMLWAADDFLEDLPFEQIG 33
 DB 398 LDPVKDVLILSALRRMLWAADDFLEDLPFEQIG 430

RESULT 3
 Q14496
 ID Q14496 PRELIMINARY; PRT; 444 AA.

AC Q14496;
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 DE Hepatitis C-associated microtubular aggregate protein p44.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=95010078; PubMed=7925411;
 RA Kitamura A., Takahashi K., Okajima A., Kitamura N.;
 RT "Induction of the human gene for p44, a hepatitis C-associated
 RT microtubular aggregate protein, by interferon-alpha/beta.";
 RL Eur. J. Biochem. 224:877-883(1994).
 DR EMBL; D28915; BAA06043.1; -
 DR EMBL; D28908; BAA06043.1; JOINED.
 DR EMBL; D28909; BAA06043.1; JOINED.
 DR EMBL; D28910; BAA06043.1; JOINED.
 DR EMBL; D28911; BAA06043.1; JOINED.
 DR EMBL; D28912; BAA06043.1; JOINED.
 DR EMBL; D28913; BAA06043.1; JOINED.
 DR EMBL; D28914; BAA06043.1; JOINED.
 DR Genew; HGNC:16938; IFI44.
 DR GO; GO:0009615; P:response to virus; TAS.
 SQ SEQUENCE 444 AA; 50502 MW; 39B7DBDA0F3F8DE2 CRC64;

Query Match 100.0%; Score 168; DB 2; Length 444;
 Best Local Similarity 100.0%; Pred. No. 1.2e-15;
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LDPVKDVLILSALRRMLWAADDFLEDLPFEQIG 33
 DB 398 LDPVKDVLILSALRRMLWAADDFLEDLPFEQIG 430

RESULT 4
 Q99984
 ID Q99984 PRELIMINARY; PRT; 413 AA.

AC Q99984;
 DT 01-MAY-1997 (TReMBLrel. 03, Created)
 DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Homo sapiens mRNA expressed in osteoblast, complete cds.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cancellous bone;
 RA Ohno I., Hashimoto J., Takaoka K., Ochi T., Okubo K., Matsubara K.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB000115; BAA19056.1; -
 SQ SEQUENCE 413 AA; 46933 MW; 026A6813174E530F CRC64;

Query Match 78.6%; Score 132; DB 2; Length 413;
 Best Local Similarity 78.8%; Pred. No. 1.8e-10;
 Matches 26; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 LDPVKDVLILSALRRMLWAADDFLEDLPFEQIG 33
 DB 371 LDPVKDVLILSALRRMLWAADDFLEDLPFEQIG 403

[illegible]

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Kryzinski M.I., Skalska U., Smallos D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16989-16993 (2002).
RN [2].
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RC TISSUE=Mammary tumor. WAP-TGF alpha model. 7 months old;
RA Strausberg R.;
RA Submitted (APR-2002) to the EMBL/GenBank/DBSJ databases.
DR EMBL; BC026901; AAH26901.1; -;
DR MGD; MGI:2443016; A430056A10Rik.
FT NON_TER 1
SQ SEQUENCE 115 AA; 13159 MW; 86AD44C8495224DD CRC64;
Query Match 63.1%; Score 106; DB 2; Length 115;
Best Local Similarity 64.5%; Pred. No. 2.8e-07;
Matches 20; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
QY 1 LDPVKDVLISALRRMLWAADDFLEDPFEQ 31
DB 85 LDPVKDMLISALKSILTYTANEFLEDPFNK 115
|||||:|||||:|:|:|||||:
|||||:|||||:|:|:|||||:
RESULT 8
Q8BV66 PRELIMINARY; PRT; 422 AA.
AC Q8BV66;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched
DE library, clone:A430056A10 product:MICROTUBULE ASSOCIATED PROTEIN 44,
DE full insert sequence.
DE Name=A430056A10Rik;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1].
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Meth. Enzymol. 303:19-44 (1999).
RN [2].
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
RN [3].
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [4].
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20499374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalization and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630 (2000).
RN [5].

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RX MEDLINE=20530913; PubMed=11078661;
RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,
RA Konno H., Akiyama J., Nishi K., Kitsuai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,
RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384 format
RT sequencing pipeline with 384 multicapillary sequencer.";
RN Genome Res. 10:1757-1771 (2000).
RN [6].
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Thymus;
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Konno H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Nunazaki R., Ohno M., Ohsato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akaiura S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBSJ databases.
RL EMBL; AK079749; BAC37740.1; -;
DR MGD; MGI:2443016; A430056A10Rik.
DR GO; GO:0015629; C:actin cytoskeleton; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; IEA.
DR InterPro; IPR002097; Profilin.
SQ SEQUENCE 422 AA; 47852 MW; FB31070ABF1D4F5E CRC64;
Query Match 63.1%; Score 106; DB 2; Length 422;
Best Local Similarity 64.5%; Pred. No. 1.1e-06;
Matches 20; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
QY 1 LDPVKDVLISALRRMLWAADDFLEDPFEQ 31
DB 392 LDPVKDMLISALKSILTYTANEFLEDPFNK 422
|||||:|||||:|:|:|||||:
|||||:|||||:|:|:|||||:
RESULT 9
Q9ER37 PRELIMINARY; PRT; 422 AA.
AC Q9ER37;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Microtubule associated protein 44.
DE Name=A430056A10Rik; Synonyms=Map44;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1].
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Dron M., Meritet J.F., Tovey M.G.;
RA Submitted (OCT-2000) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AJ299405; CAC13979.1; -;
DR MGD; MGI:2443016; A430056A10Rik.
DR GO; GO:0015629; C:actin cytoskeleton; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; IEA.
DR InterPro; IPR002097; Profilin.
SQ SEQUENCE 422 AA; 47880 MW; F267831AA0054856 CRC64;
Query Match 63.1%; Score 106; DB 2; Length 422;
Best Local Similarity 64.5%; Pred. No. 1.1e-06;